An Evaluation of Methods of Estimating Premorbid Ability.

by

John Webster Moore

D.Psychol. University of Edinburgh 1999

ACKNOWLEDGEMENTS

Thanks are due in particular to sometime clinical trainee, now colleague and friend, Professor John Crawford, University of Aberdeen for access to the databases on which this thesis is based. He was responsible for recruiting the clinicians including myself from Aberdeen, London, Stoke-on-Trent and Edinburgh for the data on clinical cases, and it was he who supervised the research team that recruited the healthy sample. He was generally a fund of encouragement, constructive criticism and support. Thanks are also due to my NHS manager and colleague Mr CT Blaiklock, Head of Neurosurgery, Aberdeen Royal Infirmary, who encouraged my application to conduct this study as part of my continuing professional development.

ABSTRACT

The detection and quantification of acquired intellectual impairment is an activity of cardinal importance in contemporary clinical neuropsychology. This is complicated by inter-individual and intra-individual variability, whereby simply comparing a patient's test performance with the relevant normal values would not provide information about any change in functioning that may have taken place.

A major aim of the present study was to evaluate regression models for the estimation of premorbid ability. The use of demographic variables, a present ability measure (the NART) and the combination of these variables was evaluated.

The vast majority of previous research on the estimation of premorbid ability has used summary IQs as the criterion variables. In the present study the ability of these methods to estimate individual WAIS-R subtest scores was evaluated. The models were tested in a large sample (N=245) of healthy participants which was broadly representative of the adult UK population in terms of the distributions of age, sex and social class. Thus this study replicates and extends previous work in the US in which demographically based regression models were developed from the WAIS-R standardisation sample.

In the present study it was found that the models combining NART and demographic variables were most successful in estimating variance in subtest performance, estimating from 30% to 71% and from 28% to 52% of the variance in verbal and non-verbal subtest performance respectively. Models estimating from

demographic information alone and from NART alone are also presented for particular clinical circumstances where either NART or demographic information is not available.

For comparative purposes models estimating WAIS-R factor scores and summary IQs from NART and demographic variables were also constructed. It was found that the models combining NART and demographic information were most successful in estimating variance in indices of WAIS-R performance. They accounted for 68%, 27% and 41% of the variance in V, PO and A/C, and for 71%, 38% and 65% of the variance in VIQ, PIQ and FSIQ.

Incorporating interactions between predictor variables into the predictor models did not lead to practically useful improvement in prediction.

To examine whether the discrepancies between estimated premorbid scores and obtained scores would successfully discriminate between healthy and impaired individuals, a heterogenous sample of neurological cases (N=298) was also employed.

Surprisingly it was found that discrimination between healthy and clinical cases was not practically enhanced using discrepancies between estimated and obtained ability scores over that which was achieved using obtained test scores alone.

The NART was shown to be impairment sensitive in a general clinical sample in line with previous research with 15% fewer clinical cases found to have estimated-obtained subtest discrepancies according to the NART models compared to those that were so identified using the demographic models.

DECLARATION

An Evaluation of Methods of Estimating Premorbid Ability.

The above thesis has been composed by me; the work of which it is a record has been done by me, and it has not been accepted in any previous application for a higher degree. All quotations have been distinguished by quotation marks and the sources of information acknowledged.

JOHN W. MOORE

September 1999.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1. The quantification of impairment.

Clinical neuropsychology is frequently concerned with the detection and quantification of impairment. There are many potential pitfalls in attempting to establish the nature and extent of potential impairment, primarily due to the considerable inter-individual and intra-individual variability in premorbid ability. Simply measuring a patient's current performance on a particular measure will not provide all one needs to know about potential impairment without reference to expectations about the likely premorbid level of functioning pertinent to that individual. This problem has exercised the wit and imagination of clinicians throughout the history of psychology, initially in the form of simple clinical judgements alone. The quest for greater precision in the determination of impairment has led to the development of a variety of approaches reviewed subsequently in this study. Of particular interest have been attempts to quantify the potential contribution of demographic variables, and regressing the criterion variable (IQ) against them. The first study incorporating this approach was by Wilson, Rosenbaum, Brown, Rourke, Whitman & Grisell (1978) who built regression equations using the *defined* demographic variables of age, sex, education, race and occupation as predictors. Using these, they were able to predict 54% of the variance in WAIS Full-Scale IQ. There have been numerous further studies replicating and extending this work subsequently, extensively reviewed below, and to which this study will make further contribution. This study is primarily concerned to evaluate the utility of using estimator

models based on demographic variables and a present ability measure in differentiating between clinical cases and normal controls.

Estimating premorbid ability -and thereby enabling the quantification of impairment in individual cases -is an activity of cardinal importance in contemporary clinical neuropsychology, whether for clinical, research, or medico-legal purposes (Crawford, 1992).

Firstly, this study will review previous work on the estimation of premorbid ability and present arguments in favour of the need to develop demographically-based estimates of premorbid ability for individual subtests of the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981). The relative contributions to prediction of demographic variables alone, demographics combined with performance on a present ability measure, and performance on a present ability measure alone will be evaluated. Secondly, in view of the superior construct validity of WAIS-R factor scores (Atkinson, 1991; Crawford, Johnson, Mychalkiw & Moore, 1997), it is proposed to evaluate factor score models for estimating premorbid ability, based on demographic information and a present ability measure, alone and in combination. It is anticipated that the results of these efforts, if successful, would be of widespread clinical utility in contemporary clinical neuropsychology in the UK with potential application elsewhere in the English-speaking world.

This chapter is concerned with a review of the use of demographic information and present ability measures in clinical practice, the measurement of general intellectual ability using the WAIS-R, (Wechsler, 1981) and a consideration of the relative merits of

different indices of intellectual ability. These indices include firstly, individual subtest scores; secondly, summary Verbal, Performance and Full-Scale IQ's (VIQ, PIQ and FSIQ), and thirdly; factor scores. Recent work by Paolo, Ryan, Troster & Hilmer (1996), presenting demographically-based regression equations to predict variance in individual WAIS-R subtest scaled scores is reviewed.

The intention in the present study is to attempt to replicate Paolo *et al.'s* (1996) findings in a healthy UK sample, and to evaluate the extent to which the models generated are powerful predictors of the criterion variables. Secondly, it is intended to establish if the inclusion of NART as a putative present ability measure produces models that are significantly more powerful predictors of the criterion variables.

Following this preliminary evaluation, it is proposed to evaluate the extent to which the combination of measures of current functioning and estimates of premorbid functioning improves our ability to discriminate between healthy and impaired samples over the discrimination achieved using measures of current functioning alone.

If the outcomes of these evaluations are positive, the models generated would be of significant utility in clinical neuropsychology, as in clinical practice, clinicians frequently require to make inferences from *individual* WAIS-R subtest scores (e.g., McKinlay & Gray, 1992).

1.2. The rationale of deficit measurement.

It has been argued that normative comparison standards for assessing deficit are only appropriate when the function, skill or capacity in question is exhibited by all normal individuals, irrespective of general intellectual ability or personal experience (Lezak, 1983). In this situation, direct measurement of deficit is possible. However, where the function, skill or capacity is *normally distributed* in the general population, normative comparison standards will be unhelpful in detecting whether or not a deficit with respect to the previous state is present.

Franzen, Burgess & Smith-Seemiller (1997) note the recent development of more sophisticated *normative* information, taking account of variables including age, sex and education, but because of the large number of data cells in such studies, the samples must be of great size. For example, Heaton, Chelune, Talley, Kay and Curtiss (1993) presented normative data from 486 subjects, broken down into two levels for sex, six levels for education, and ten levels for age, whereby the *average* cell size is only four subjects, to represent the entire sample range within that cell. Also when subjects change age group, there can be a stepwise change in the normal values suddenly appropriate for the interpretation of their test scores. With this approach, conducting meaningful clinical interpretation becomes logistically daunting, and some more practical alternative for evaluating the performance of individual patients is required.

Direct measurement of deficit in respect of a *distributed* function, skill or capacity is occasionally possible where premorbid test scores exist, and can be obtained. For example, Canter (1951) compared the current test scores of Army veterans with

multiple sclerosis with their premorbid scores (obtained on induction to the service) on the same battery of tests. However, as Lezak (1983) has noted, in practice, such data is rarely in existence or easy to obtain by busy clinicians with finite resources of time and funding. Contemporary researchers are also likely to be interested in scores on measures that were not developed when the premorbid assessments were conducted. The latter may reflect advances in the theoretical understanding of the cognitive system and the development of measures designed to assess these new theoretical constructs. An additional problem is that general ability in populations appears to improve over time, which has led to the requirement, for example, to restandardise the Wechsler Intelligence Scales. As Crawford (1992) has discussed, the replacement of the WAIS by the WAIS-R was justified by the reliable observation that the WAIS was inflating the mean IQ in the general population by around half a standard deviation. It has also been observed that IQ gains are not uniform across all cognitive measures. This further complicates the interpretation of discrepancies between test results where, for example, clinicians may be interested in measuring subtest scatter; one may no longer be able to assume that the population mean scores on Wechsler subtests are equivalent, and that there may be inbuilt discrepancies between subtests simply as a result of differential IQ gains in the general population.

As Lezak (1995) has discussed, there are substantial inter-individual differences in intellectual ability reflected in a distribution of scores obtained when a test is applied to measure an intellectual variable in any general population sample. As noted, quantifying impairment in this situation is problematic; simply comparing a patient's

current test performance with the relevant test norms would be confounded by inter-individual variability whereby, for example, a test score that represents significant impairment for one individual might represent normal performance for another of lesser premorbid ability. Change in level of functioning might have serious implications for return to previous lifestyle or occupation, so that *relative* change may be of greater significance than *absolute* level, in the individual case. Because of this, *normative* comparison standards are only useful in dichotomous situations, and in most neuropsychological assessment we must supplement these with *individualised* comparison standards when attempting to quantify acquired impairment. Hence the need for a means of *estimating* premorbid ability. To this end, predictor variables (often incorporated in regression equations) are widely employed in clinical neuropsychology as an alternative to conventional normative data.

1.3. Estimating premorbid ability.

There are three main approaches currently employed to estimate premorbid ability:

(A) Estimates based on a present ability measure, which is thought to be relatively resistant to the effects of CNS impairment;

(B.) Estimates based on demographic variables which are known to be correlated with IQ or other index of intellectual ability;and,

(C).Estimates based on demographic information and a present ability measure in combination.

Each of these approaches has relative strengths and weaknesses, and each embodies particular assumptions. In particular clinical circumstances, one approach may be more appropriate, and any particular approach may be absolutely inappropriate. Each of these approaches will be reviewed in turn.

1.3.1. Approach (A): Estimates based on present ability measures.

This approach involves using a present ability measure, and depends upon two important assumptions. Firstly, it is assumed that a particular present ability measure of an individual's cognitive performance will predict a reliable estimate of their likely performance on another measure. As O'Carroll (1995) has discussed, throughout the history of psychology it has been reliably observed that, in healthy subjects, correlations between any two cognitive measures are generally positive. This assumption, then, seems reasonable. Secondly, it is assumed that not all tests of cognitive ability will be equally affected by injury or disease, and that some abilities will be unaffected. This state of affairs reflects the differential vulnerability of some cognitive abilities to the effects of injury or disease and to variations in the severity and distribution of the pathology in the individual case.

"Hold-Don't Hold" Methods.

Historically, one approach to the determination of intellectual deterioration was via a comparison of so-called "hold"(e.g. Vocabulary), versus "non-hold" (e.g. Block Design) subtests of the Wechsler scales (Coolidge, Peters, Brown, Harsch & Crookes, 1985). However, it is well known that even the best "hold" subtests are vulnerable to the effects of dementia (Hart, Smith & Swash, 1986; Crawford, Besson & Parker, 1988; Sharpe & O'Carroll, 1991). This followed work by Wechsler (1958), who proposed a Deterioration Quotient derived by subtracting age scale scores for "don't hold" tests from those pertaining to "hold" tests, and a number of variations of this procedure have been described. Yates (1956), advocated simply using Vocabulary as an index of premorbid ability because of its high correlation with FSIQ.

The foregoing approach leads logically on to the development of models to differentiate particular neurological conditions from controls. An example of this is the Fuld Cholinergic Profile (Fuld 1982; 1984), which was proposed to identify patients with possible Alzheimer's Disease (DAT). Fuld (1984) initially reported a 44% sensitivity and 96% specificity of the model in 12 patients with DAT relative to 28 patients with other dementia syndromes. However, Brinkman & Braun (1984) applied the model in a sample of 23 patients with DAT and a group of 39 patients with multi-infarct dementia. The model was found to have a correct classification rate of 57% with 94% specificity. Unfortunately, Goldman, Axelrod Tandon & Berent (1993) reported that 15% of a sample of acutely-ill schizophrenics not on anticholinergic medication exhibited the profile, and further studies have shown that the pattern is present in normal

elderly subjects (Logsdon, Teri, Williams, Vitiello & Prinz, 1989), closed head injury patients (Heinricks & Celinski, 1987), and in depressed patients (Bornstein, Termeer, Longbrake, Heger & North, 1989). A review of 18 studies employing the Fuld model by Massman & Bigler (1993), showed overall sensitivity 24.1% and specificity in clinical samples of 88.5%, suggesting that the model is more limited than was first assumed. Furthermore, Obonsawin, Robertson, Crawford, Perera, Walker, Blackmore, Parker & Besson (1998) showed in a prospective study that cholinergic blockade achieved by administering scopolamine to 12 healthy subjects did not lead to the characteristic WAIS-R profile reported by Fuld (1984).

All of these approaches must be evaluated against a clear appreciation of the base rates of subtest scatter in the general population. The fundamental problem with the approach, however, is simply the well-known vulnerability of the so-called "hold" tests in the face of neurological disease or injury (Matarazzo & Prifitera, 1989).

The Best Performance Method.

This approach to estimating premorbid ability is a clinical method which may include an examination of test information, observed and reported behaviour, and evidence of premorbid achievements to assist in a clinical estimation (Lezak, 1995). Running somewhat counter to her arguments against the use of summary IQ measures (Lezak, 1988a), an underlying assumption of the best performance method is that the average individual's test scores will group round some hypothetical mean. In support of this notion, Lezak (1995) lists research on the general ability factor *g* as supporting the

validity of this assumption. However, Mortensen, Gade & Reinisch (1991) note considerable intra-individual scatter in the test scores of healthy subjects, and in spite of variance attributable to *g* they found that the best performance method overestimated premorbid ability in healthy adults and in patients with cerebral atrophy. Matarazzo & Prifitera (1989) have shown considerable subtest scatter in normal subjects comprising the WAIS-R standardisation sample. These authors demonstrated that the *mean* difference between highest and lowest subtest scaled scores was 6.66. Approximately 30% showed differences of 8 or more points. Furthermore, subjects deviating extremely from mean IQ had higher subtest differences. These authors argue that single subtest scores in healthy subjects would produce widely dispersed estimates of FSIQ. Crucial to the interpretation of subtest scatter is the distinction between *reliable* and *abnormal* differences (Crawford, 1992) which is discussed further in ensuing sections. A further problem for the best performance method is that when test scores on diverse tests are considered, they do not have the advantage of the Wechsler scales whereby they are standardised on the same subjects. In these circumstances, the clinician has the added problem of mapping between different standardisation samples.

As an alternative to using a patient's best score on a test battery as an index of their likely premorbid level of functioning, attention has more recently been focused on specific measures of present ability which have been shown to be relatively robust in the face of brain injury or disease. A number of potential measures are discussed in the ensuing section, concentrating principally on the National Adult Reading Test (NART; Nelson, 1982).

The National Adult Reading Test.

As O`Carroll (1995) has noted, The National Adult Reading Test (NART, Nelson, 1982) has become extremely popular in clinical practice and research as a quick and simple means of obtaining an estimated measure of premorbid ability. Its psychometric properties have been extensively reviewed by Crawford (1992), O`Carroll (1995) and by Franzen, Burgess & Smith-Seemiller (1997), and this review draws upon these sources.

As Crawford (1992) has argued, any current ability measure must fulfil the following four criteria if it is to qualify as a valid means of estimating premorbid ability: Firstly, it must have adequate reliability; secondly, it must correlate highly with IQ in the general population (criterion validity); thirdly, performance on the measure must be highly resistant to the effects of CNS injury or disease (robustness); and fourthly, it must improve our ability to detect impairment over the use of impairment-sensitive measures alone.

The development of the NART followed on from previous work using the Schonell Graded Word Reading Test (GWRT; Schonell, 1942). The GWRT had been developed to assess reading ability in children, but it had a ceiling at IQ 115, restricting its applicability in the general population. Nelson and McKenna (1975) developed the NART with a wider range of applicability. The NART is a new reading device based, as is the GWRT, on the observation that oral reading ability (accuracy of oral

pronunciation) is relatively resistant to deterioration in patients with dementia. In distinction to the GWRT, the NART comprises only irregular words. In the general population, the level of reading ability is highly correlated with general intellectual ability, and once it is established, reading ability is highly resistant to deterioration as other aspects of cognitive ability diminish.

The NART consists of a list of 50 short *irregular* words; that is, they do not follow normal grapheme-phoneme correspondence rules (e.g., chord, radix, placebo etc.,). Because of this characteristic, subjects should not be able to provide the correct pronunciation unless the particular word formed part of their vocabulary prior to the onset of dementia or brain insult. Nelson and O`Connell (1978) have argued that the test requires previous familiarity with the stimulus words and that it makes little demands on current cognitive capacity. They reported that reading ability of demented patients was unimpaired relative to controls, and it has been concluded that the NART provides a valid estimate of premorbid ability in patients with dementia. The NART's criterion validity was investigated by Crawford, Stewart, Parker, Besson & De Lacey (1989), reporting that NART predicted 66%, 72% and 33% of the variance in WAIS FSIQ, VIQ and PIQ respectively, and that it is therefore a relatively powerful predictor of WAIS FSIQ and VIQ, whilst it is relatively poor at predicting PIQ. Table 1 summarises a number of studies relevant to the assessment of the criterion validity of the NART in UK, US, Canadian, and Australian samples. The various studies are generally consistent in terms of the ability of the reading measure employed to estimate variance in IQ with

the exception of the Australian study, where the NART could account for only 26% of the variance in WAIS-R IQ. Willshire, Kinsella & Prior (1991) found that NART was a powerful predictor of WAIS-R IQ in middle age to old subjects but not in young Australians, possibly reflecting fundamental changes in the educational system.

Evidence of its construct validity was provided in a factor-analytic study (Crawford, Stewart, Cochrane, Parker & Besson, 1989), where the NART loaded 0.85 on *g*, the first unrotated principal component using NART and WAIS data. The determination of it's construct validity can only be ascertained in a *normal* sample, where performance on the measures concerned is unaffected by potential illness factors.

The NART is one of the most reliable tests in clinical practice (Crawford, 1989), in that it has high split-half reliability (Crawford, Stewart, Parker, Besson & De Lacey, 1989; Nelson 1982), high inter-rater reliability (Crawford 1992; O`Carroll 1987), and high test-retest reliability (Crawford *et al.*, (1989).

In summary, with regard to its psychometric properties, the NART has impressive reliability and criterion validity, but the further key issue concerns the other component of its validity- (robustness)- that is to say does the pronunciation of irregular words survive in conditions in which other aspects of cognitive ability are impaired?

Table 1.1: Criterion validity of the NART.

Study	Country	N	Test	Criterion	FSIQ Var.
Nelson (1982)	UK	120	NART	WAIS	55%
Crawford *et al.* (1989)	UK	151	NART	WAIS	66%
Blair & Spreen (1989)	Can	66	AM-NART	WAIS-R	56%
Crawford (1990)	UK	200	NART-R	WAIS-R	59%
Willshire, Kinsella & Prior (1991)	Aust	104	NART	WAIS-R	26%
Nelson & Willison (1991)	UK	181	NART	WAIS-R	72%

The NART in neurological and psychiatric conditions.

Studies of the performance of the NART in a variety of neurological and psychiatric conditions have been extensively reviewed by Crawford (1992), O'Carroll (1995) and Franzen, Burgess & Smith-Seemiller (1997), and selected studies are summarised by diagnosis below. Although the present work is concerned primarily with ability measures and predictors in normal and neurological samples, the behaviour of these measures in psychiatric conditions is of considerable importance in clinical

neuropsychological practice and research. This reflects the issue of the differential diagnosis of depressive illness and dementia,- described by Lezak, (1983, p.234), as "probably the knottiest problem of differential diagnosis"- and the frequency of co-morbidity of mental illness and neurological disease. A more detailed review of the performance of ability measures in mental illnesses may be found in Crawford (1992), O`Carroll (1995) and Franzen, Burgess & Smith-Seemiller (1997).

It is worth considering the potential methodological routes to investigating the behaviour of irregular word reading in subjects with neurological diseases and psychiatric illnesses. Firstly, some studies compare the performance of cases to controls matched on all other potentially contributory variables in cross-sectional designs. Secondly, longitudinal studies have been attempted in degenerative conditions with serial assessments, with each subject as their own comparison. Thirdly, other studies have investigated the relationship of NART scores with scores on dementia-sensitive measures, with patients stratified according to scores on the dementia measure.

Alzheimer`s disease

Alzheimer`s disease accounts for approximately one half of all cases of dementia and is thought to affect between 5 and 6 percent of individuals over age 65 as well as those affected below 65 years of age. It is a diagnosis only verified post-mortem, unless brain biopsy is carried out in life. Histo-pathologically it is characterised by the presence of senile plaques and neurofibrillary tangles in the neocortex and hippocampus (Tomlinson, Irving & Blessed, 1981). These alterations in the integrity of brain are

associated with mental deterioration, which runs a progressive course. The earliest signs of the disease are usually comprised of a disturbance of new additions to long term memory, and dysthymic disturbance in the form of depression and irritability, although occasionally a seizure or language disturbance heralds the onset of the disease (Lezak, 1983). Multi-infarct dementia may be clinically indistinguishable from Alzheimer`s disease (Lezak, 1983; Walton, 1977), but is often characterised by a `stepwise` progression (Hachinski, Linnette, Zilhla, Du Boulay, McAllister, Marshall, Ross Russell & Symon, 1975) and may be associated with multi-level vascular disease in the heart, aorta and carotid arteries, distally in the legs, as well as in the brain.

A number of studies have reported no apparent detrimental effect on reading ability in early Alzheimer`disease and multi-infarct dementia (Nebes, Martin & Horn, 1984; Cummings, Houlihan & Hill, 1986; O`Carroll and Gilleard, 1986; Crawford, Parker & Besson, 1988). A further important study by O`Carroll, Baikie & Whittick, (1987) assessed patients with Alzheimer`s disease and multi-infarct dementia on two occasions with assessments separated by one year and reported no decline in reading ability in the face of progressive neurological disease. This is an important study in that longitudinal studies permit the stability of measures to be examined in progressive disease. However, Fromm, Holland, Nebes & Oakley (1991) tested patients with Alzheimer`s disease annually for three years and showed that NART performance deteriorated and that NART scores were correlated with dementia severity. Similarly, Patterson, Graham & Hodges (1994), assessed 45 patients with Alzheimer`s disease and

found that NART performance was correlated with MMSE dementia severity. They suggest that NART performance underestimates premorbid ability by approximately 15 IQ points when patients are moderately impaired.

Stebbins, Gilley, Wilson, Bernard & Fox (1990) have shown that NART performance is impaired in patients with language disturbance, and must therefore be used with caution in such patients.

In summary, with respect to Alzheimer`s disease and multi-infarct dementia, whilst there is impressive evidence in favour of the NART`s reliability and validity in normal subjects, and it is capable of predicting impairment in early Alzheimer's Disease, the performance of patients is not maintained as the disease becomes more advanced.

Schmand, Geerlings, Jonker & Lindeboom (1998) reported the six-year stability (robustness) of the Dutch version of the NART (DART) in a sample of 197 subjects, normal at baseline but thought to be possibly dementing, compared to a cognitively intact control group. They examined changes in DART score relative to severity and change in MMSE score. Their findings were consistent with those of Patterson *et al.* (1994), and Taylor, Salmon, Rice, Bondi, Hill, Ernesto & Butters (1996), noting that DART underestimates premorbid IQ by as much as one standard deviation in patients with moderate to severe dementia (MMSE<14). In the early stages of dementing illness, when the diagnostic uncertainty is greatest, DART scores declined at only a modest rate relative to controls.

A proportion of patients with these conditions have language disturbance sometimes as the presenting feature, and, as noted, Stebbins *et al.*,(1990) have shown that NART performance does not provide an accurate predicted ability level in these patients. Again, the discrepancy between predicted and obtained NART should identify such patients for further assessment with language screening instruments, using the regression equation provided by Crawford, Allan, Cochrane & Parker (1990), which predicts a patient's NART performance from background demographic information.

Korsakoff`s syndrome

Korsakoff`s syndrome typically affects alcoholics and reflects damage to medial temporal lobe structures in the limbic system caused by thiamine deficiency (vitamin B1). The condition is manifest by an amnesic state whereby new information cannot be consolidated, and therefore neither retrieved nor utilised. It is often associated with a propensity to confabulate. Because of the relatively circumscribed nature of the neurological impairment in Korsakoff`s syndrome, and the observation that general intellectual ability is unimpaired in this condition it is perhaps surprising that Crawford *et al.*, (1988) showed that their sample of 12 alcoholic Korsakoff patients performed significantly more poorly than individually matched controls on NART. This finding was subsequently replicated by O`Carroll, Moffoot, Ebmeier & Goodwin (1992). O`Carroll (1995) suggests that the possible explanation is that these patients may not error-check

their performance, tend to read phonetically in a stimulus-bound manner, and that this may represent a component of confabulation syndrome.

Huntington`s disease

Huntington`s disease (HD) is a progressive, autosomally dominant, progressive neurodegenerative disease with complete lifetime penetrance. The most prominent neuropathological features are neuronal loss and gliosis in the striatum, particularly in the caudate nucleii, with prominent dysexecutive features reflecting early disruption of fronto-striatal pathways (Caine, Hunt, Weingartner & Ebert, 1978; Lezak, 1995; Starkstein, Brandt, Folstein, Strauss, Berthier, Pearlson, Wong, McDonnell & Folstein, 1988). The behavioural syndrome is characterised by involuntary choreiform movements, cognitive impairment of a progressive nature and changes in personality and affect (Brandt, 1991; Lishman, 1987). As part of a large study comparing a diverse clinical sample with matched controls, Crawford, Parker & Besson (1988) reported that six HD patients performed poorly on NART with respect to matched controls. This finding was confirmed by Blackmore, Crawford & Simpson (1994) in a larger study. These authors concluded that using demographic variables might be a more suitable method for estimating premorbid ability in this condition.

Traumatic brain injury

Crawford, Parker & Besson (1988), reported on a series of closed head injury patients (N=19), as part of a large multi-pathology sample of neurological cases who

exhibited evidence of intellectual and social impairment and showed that there was no difference between NART and vocabulary estimated IQ using Nelson and McKenna's (1975) regression equation to convert WAIS Vocabulary age-graded scaled scores to estimated Full Scale IQs. This pathological group was distinguished from the other neurological groups in the study in that they did not differ from matched controls on WAIS Vocabulary. Simons (1997) reported that NART performance was unimpaired in a sample of 19 severe head-injury subjects despite significant deficit on WAIS-R FSIQ. This finding was confirmed in a further study by Watt & O'Carroll (1999) comparing NART performance in a group of 25 head injury patients with that of a group of 50 healthy controls who were significantly different on current intellectual functioning.

Idiopathic Parkinson's Disease

Idiopathic Parkinson's Disease (IPD) is a movement disorder characterised by rigidity, tremor and bradykinesia, the so-called triad of motor symptoms. These symptoms arise from a loss of dopaminergic innervation of the striatum, which in turn is the result of degeneration of the substantia nigra. In addition to this, it has been observed that there is dysfunction in the meso-cortico-limbic dopamine system. Scatton, Rouquier, Javoy-Agid & Agid (1982) reported reduced levels of dopamine in the cortex generally, and in the hippocampus, and Javoy-Agid & Agid (1980), described cell loss in the ventral tegmental area, which contains the cells of origin of the meso-cortico-limbic projection.

James Parkinson, in his original description of the condition stated that "the senses and intellect" were "uninjured" (Quinn, Brown & Marsden, 1986), but it is now clear that a proportion of patients with IPD have widespread dysfunction in other neurotransmitter systems. These include the cholinergic, noradrenergic, seretonergic and GABAergic systems (Crawford, Besson & Ebmeier, 1990). In these patients, cognitive impairment has been shown, and it would be surprising if this were not the case, in view of the widespread neurotransmitter system dysfunction. The magnitude of these deficits constitutes dementia in a proportion of cases, and in non-demented cases, cognitive deficits pertaining to the frontal lobes have been shown (Levin, Llabre & Weiner, 1989). These authors showed that deficits were observable at an early stage of the disease, and clearly a measure capable of providing a valid estimate of premorbid ability in this condition would be of considerable clinical utility. Crawford (1990) examined the stability of the NART in IPD, and examined correlations between NART performance and measures of illness severity. A series of 60 patients were examined using the cognitive measures of the MMSE and NART, and the demographic variables of age, sex, years of education, and years since illness onset were recorded. Clinical severity of physical disability in IPD was evaluated using the Hoehn & Yahr Scale (Hoehn & Yahr, 1967), and the severity of tremor, rigidity and bradykinesia were evaluated to provide a composite measure of motor symptoms using the Webster Scale (Webster, 1968). The correlations between NART and the two measures of severity of physical disability did not reach significance, and the NART was not found to be impaired in IPD relative to controls. When the sample was split into two groups on the basis of MMSE score,

Crawford (1990) concluded that the NART was mildly impaired in the IPD demented subgroup, and unimpaired in the non-demented patients versus matched controls. Crawford concluded that NART performance was unimpaired in IPD, and could validly be used in this condition as a means of estimating premorbid ability.

Glioma and whole brain irradiation

16 patients with cerebral glioma who had received whole-brain radiotherapy were compared with matched controls (Ebmeier, Booker, Cull, Gregor, Goodwin & O'Carroll 1993) and, after controlling for demographic variables, it was found that patients made more errors on NART than controls. In addition, using the Crawford, Allan, Cochrane & Parker (1990) regression equation to predict NART errors from demographic variables, patients made more errors than predicted. The measure must therefore be used with caution in this clinical group.

The NART in depressive illness.

As previously noted, the differential diagnosis of dementia versus depression has been described as "probably the knottiest problem of differential diagnosis" (Lezak, 1983, p.234) and such patients are commonly sent for neuropsychological evaluation, in view of the equivocal nature of their clinical presentation. Crawford, Besson, Parker, Sutherland & Keen (1987) reported that the NART performance of a sample of depressed patients did not differ from that of *matched controls.* It is worth noting that

performance on the WAIS Vocabulary subtest did not hold in this group of patients. On conventional neuropsychological tests, these patients may appear mentally impaired, but they do not appear to exhibit the decline in NART performance exhibited by the patients reported by Schmand, Geerlings, Jonker & Lindeboom (1998) with moderate to severe dementia (MMSE<14). However, as O'Carroll (1995) has indicated, comparison of unequivocally depressed, and unequivocally normal cases, is not representative of the typical clinical situation in which the differentiation is attempted. Because disturbance of memory function is the cardinal presenting clinical feature of the dementias, Schlosser & Ivison (1989) proposed that a comparison between NART and memory indices (in contrast to intellectual ability) might permit clearer differentiation between depressed and demented patients. In fact, the attempt to differentiate on the basis of assessment of memory might be even more difficult since memory impairment is a prominent manifestation of depressive pseudo-dementia, unless, of course, it can be shown that disturbance of memory in the two groups is qualitatively different.

O'Carroll, Curran, Ross, Murray, Riddle, Moffoot, Ebmeier & Goodwin (1994) reported a discrepancy analysis involving the NART and the Wechsler Memory Scale-Revised (WMS-R, Wechsler, 1987) comparing the performance of three groups of patients, more closely representing the clinical situation than the conventional patient-and-normal-control design. They compared patients with probable Alzheimer's disease, patients with clinical major depression and a group of matched controls to see if these groups could be differentiated on the basis of memory impairment. Whilst group mean differences were observed, there was still overlap between groups. Thus, they could not

establish any means of reliable, categorical differentiation for use in the clinical context when confronted with the differential diagnosis of an individual case. Furthermore, there was no group of clinically equivocal patients in their study. These patients are most commonly referred for diagnostic assessment rather than those cases in which the clinical diagnosis is more clear.

The foregoing study treats dementia and depression as though they were mutually exclusive entities; that is to say, that patients cannot be suffering from both. Future research in this area may benefit from approaches which consider the *relative* probabilities of *all* possible diagnoses as in The Revised Kendrick Battery (Kendrick, Gibson & Moyes, 1979) which employs Bayes' theorem to assign *a-priori* probabilities to all combinations of diagnosis with assessments separated in time. The initial probabilities are subjected to *a-posteriori* revision using Bayes theorem in the light of results obtained from the second assessment.

The NART in Schizophrenia.

Sample characteristics have been shown to be critical in studies assessing the validity of the NART in schizophrenia. Crawford, Besson, Bremner, Ebmeier, Cochrane & Kirkwood (1992) compared the performance of two schizophrenic samples with each other, and with matched controls. One group comprised long-stay hospitalised patients, and the other comprised patients living in the community. NART estimated IQ did not differ between community resident schizophrenics and controls, but NART estimated IQ

in the hospitalised patients was significantly lower. Crawford *et al.* concluded that low NART scores could be a valid reflection of low premorbid IQ, but the hospitalised patients had been ill for longer, and may have been exhibiting a greater disease burden than the community residents. In addition, the possible effects of long-term neuroleptic use in this sample were not known. Crawford *et al.'s* study suggests that the use of the NART to obtain an estimate of premorbid IQ in acutely ill, testable patients may be justified, but that it's exclusive use with more chronic patients is inadvisable. O'Carroll, Walker, Dunan, Murray, Blackwood, Ebmeier & Goodwin (1992) also showed that a group of acutely ill, unmedicated patients with schizophrenia did not differ from matched controls with respect to NART performance.

As we have seen, it will not always be clear whether or not the NART can provide a *valid* estimate of premorbid ability in the individual case. To assess this, Crawford, Allan, Cochrane & Parker (1990), developed a regression equation to predict a patient`s NART performance from background demographic details. This provides an important means whereby it is possible to check if the instrument is underestimating premorbid ability in the individual case. Although this relative lack of robustness is disappointing, as Crawford (1992) has discussed, by the time this is likely to be a factor in an individual patient`s performance, the diagnosis should not be in doubt, thus *obviating* the need for a comparison standard to aid in it's detection. Indeed, given the scale of clinical disturbance in dementia, it would be surprising if word reading ability were *entirely* invulnerable to it's effects.

Other present ability measures.

Beardsall & Huppert (1994) were struck by the high incidence of mispronunciations of common words in their sample of well-educated normal subjects and concluded that, although it was likely that they used these words in everyday life, encountering them out of context was the likely cause of the mispronunciations, not unfamiliarity. They argued that these subjects would by definition, possess a prior correct lexical entry. On this basis they developed the Cambridge Contextual Reading Test (CCRT), where NART words are presented in an appropriate semantic and syntactic context. An example of such an approach might be: Cromwell sought to *deny* the *heir* his right to the throne. Using this material, normal subjects, poor readers and patients with dementia improved their reading scores compared to the NART test, where the words e.g., *deny* and *heir* are presented in a word list. Good readers did not benefit from the provision of contextual cues. Interestingly, the patients with dementia benefitted most of all. Further work with this promising instrument is eagerly anticipated, which might include validation against the WAIS-R, and the Wechsler Adult Intelligence Scale-III (WAIS-III, Wechsler, 1997a) and, crucially, the assessment of performance in clinical groups in whom conventional NART performance has been shown to decline. Simons (1997) found that CCRT performance did not deteriorate in a series of 19 patients with very severe brain injury, but it must be noted that NART performance was also unimpaired in the sample. Conway & O'Carroll (1997) showed that in a group of 30

patients with probable Alzheimer's disease, CCRT performance was better than NART and although the latter was correlated with dementia severity, the former was not.

Following on from the development of this contextual reading task, Baddeley, Emslie & Nimmo-Smith (1993) have developed the Spot-The-Word Test (STW), a test of lexical decision making. Subjects are presented with pairs of stimuli, one of which is a true word, whilst the other is an invented non-word. They argue that lexical decision making can be based on any one of a number of characteristics of a word, including meaning, orthographic appearance, sound, or it's general familiarity. They argue that there are a number of parallel routes to effective task performance, and it is therefore likely that performance on the task would be more resistant to the effects of brain failure.

Simons (1997) showed that STW performance did not deteriorate in his sample of 19 brain injury subjects who were given the WAIS-R, NART, CCRT and STW. Watt & O'Carroll (1999) replicated this finding in a sample of 25 closed head injury patients relative to 50 healthy controls.

Again, further data in connection with this promising development is eagerly anticipated, as with the CCRT above, particularly, as noted, in conditions in which NART has been shown to deteriorate.

1.3.2. Approach (B): Estimates based on demographic variables.

Prior to the formal use of demographic information in psychometrics, clinicians would use clinical judgement in comparing a patient's test performance against their

own expectations of an individual with a similar background. This has been a significant source of error in clinical interpretation, reflecting in part, clinicians' variable and usually limited experience of normal subjects (see e.g., Kareken, 1997, for a detailed review).

As noted, the first regression equations predicting intellectual performance from demographic information were presented by Wilson, Rosenbaum, Brown, Rourke, Whitman & Grisell (1978). These were found to predict 54%, 53%, and 42% of the variance in WAIS FSIQ, VIQ and PIQ respectively. Barona, Reynolds & Chastain (1984) subsequently presented regression equations for use with the WAIS-R in the US. This approach takes advantage of the well-established relationship between particular variables and measures of intellectual ability (e.g., Matarazzo, 1972).

In the UK, Crawford, Stewart, Cochrane, Foulds, Besson & Parker (1989) reported that the following demographic regression equation predicted 50% of the variance in WAIS Full-Scale IQ (FSIQ):

FSIQ = 104.12 - 4.38 x class + 0.23 x age + 1.36 x years education - 4.7 x sex.

However, 50% of the variance in IQ remains unexplained, but the method is superior to clinical guesswork (Kareken & Williams, 1994). Additional equations provided by Crawford *et al.* (1989) predicted 50% of the variance in VIQ, and 30% of the variance in PIQ. Crawford & Allan (1997) presented further equations for use with the WAIS-R based on a sample of 200 cases, and were able to predict 53%, 53% and

32% of the variance in FSIQ, VIQ, and PIQ. They also presented tables for evaluating predicted-obtained discrepancies in the individual case based on the base rate of occurrence of these in their sample.

The general approach is based on the assumption that demographic factors operate equally for all subjects. However, clearly, peculiar social or health factors may distort demographic experiences in an idiosyncratic way, so that their use would be manifestly unreliable. For example, severe chronic illnesses in childhood may adversely affect education and employment opportunities. If such illnesses occur early in life, the concept of *premorbid* becomes meaningless. Similarly, discrimination on grounds of race or sex, for example, may operate to disadvantage individuals relative to their true potential. This approach cannot be expected to produce an accurate estimate of premorbid ability in circumstances where the underlying assumptions with regard to demographic variables cannot be met. These factors would be expected to lead to an underestimation of premorbid ability. Conversely, affluent, well-connected parents might preferentially educate their offspring and use their personal contacts to place their children in better jobs than they might merit, leading to an overestimation of ability in such cases. Clinicians should beware of using the method mechanically without regard to background circumstances. Equally, research on the systematic evaluation of demographic predictor variables was motivated, as noted, by the wish to improve on the inaccuracy of prediction inherent in the pure clinical method whereby background circumstances are fully taken into account (see e.g. Kareken & Williams, 1994).

This approach has the advantage, however, of being completely independent of current intellectual functioning, which may be an important consideration in an individual clinical context, in subjects unable to communicate normally, as in for example clinical dysphasia, or where there are other grounds for rejecting the use of present ability measures.

Table 1.2 summarises a number of studies in which demographic variables were employed to predict variance in WAIS and WAIS-R IQ.

Table 1.2: Percentage of FSIQ variance predicted by demographic variables in the US and UK

Study	Test	Country	% variance
Wilson *et al.* (1978)	WAIS	US	54
Barona *et al.* (1984)	WAIS-R	US	36
Barona & Chastain (1986)	WAIS-R	US	43
Crawford *et al* (1989)	WAIS	UK	50
Crawford & Allan (1997)	WAIS-R	UK	53

Paolo, Ryan, Troster & Hilmer (1996) have presented demographically based regression equations to predict variance in *individual* WAIS-R subtest scaled scores using the US WAIS-R standardisation sample. This is of considerable interest to clinicians who, for

whatever reason, may wish to make inferences from one or two WAIS-R subtest scores. A number of issues pertinent to this are discussed in more detail in section *1.5.2.*

1.3.3. Approach (C): Combining demographics and present ability measures.

It has been shown that both demographic information and present ability measures can be used to make useful estimates of intellectual ability. The key issue now is to review studies in which they have been used in combination.

This approach involves building regression equations combining demographic information together with a present ability measure. In a study of 151 normal subjects, Crawford, Stewart, Parker, Besson & Cochrane (1989) built a multiple regression equation to predict WAIS IQ. Although there is considerable covariance between NART and demographic variables, most so for education and social class, combining these variables did lead to an increased prediction of variance in WAIS IQ. NART plus demographics predicted 73%, 78% and 39% of the variance in WAIS FSIQ VIQ and PIQ respectively, compared to 66%, 72% and 33% for NART alone, and 50%, 50% and 30% for demographics alone. This provides justification for the intention here to include NART with demographic information in extending the Paolo *et al.,* (1996) study, referred to earlier.

As an example of studies basing prediction on current ability measures other than NART, Krull, Scott & Sherer (1995) combined demographic information and

performance on the WAIS-R subtests of Vocabulary and Picture Completion to predict IQ values in a cross-validation study utilising the WAIS-R standardisation sample, and the problems of restricted range in predicted scores seen with the pure demographic approach (e.g., Barona, Reynolds & Chastain, 1984), were not seen. Similarly, Vanderploeg & Schinka (1995) reported a further variation of this approach including current WAIS-R performance. The fundamental limitation with these studies, however, is that it is well established that performance on WAIS-R subtests is affected by a wide range of neurological and psychiatric conditions.

1.4. Regression models for estimating premorbid performance on specific neuropsychological tests.

Whilst the principal use of the NART has been in relation to predicting premorbid WAIS-R IQ, it has also been used to estimate performance on other neuropsychological tests on it's own and in conjunction with demographic variables. For example, van den Broek & Bradshaw (1994) presented multiple regression equations for estimating premorbid Raven's Progressive Matrices scores on the basis of age and NART score.

Crawford, Moore & Cameron (1992) built a regression equation to estimate premorbid performance on the FAS verbal fluency test. They reported that in a healthy sample of (N=142) subjects, NART and verbal fluency (FAS) were highly correlated (r = .67) indicating that premorbid ability should be taken into account when interpreting verbal fluency performance. This is consistent with the report by Miller (1984), that

verbal fluency was correlated with VIQ in a sample (N=36) of healthy subjects. Furthermore, Borkowski, Benton & Spreen (1967) have reported that brain injured subjects of above average IQ obtained higher morbid verbal fluency scores than below average IQ healthy control subjects. The regression equation to predict premorbid NART derived from the healthy sample was then used to estimate premorbid NART in a general neurological sample (N=38), in which concussional brain injury and subarachnoid haemorrhage predominated. There was a highly significant difference in favour of predicted verbal fluency in the clinical sample. These authors presented a table for converting NART errors to predicted verbal fluency score.

Crawford, Obonsawin & Allan (1998) have built a regression equation which combines NART score and age to estimate premorbid scores on the Paced Auditory Serial Addition Test (PASAT; Gronwall, 1977). This is a particularly useful equation in view of the sensitivity of the measure to impairment.

As noted, Schlosser & Ivison (1989) proposed a regression equation incorporating NART and age to estimate performance on the Wechsler Memory Scale. These authors reported high success in the detection and quantification of memory impairment in patients with probable Alzheimer's dementia. However, O'Carroll, Curran, Ross, Murray, Riddle, Moffoot, Ebmeier & Goodwin (1994) showed that in a methodological sophistication of the previous work, using a group of depressed patients as well as the other two groups comprising patients with probable Alzheimer's disease and controls, that there was overlap of group distributions of scores, in spite of well

separated mean scores. None of the NART/WMS-R discrepancy quotients, which they evaluated, were usable in clinical practice for unequivocal differentiation.

As noted previously, for situations in which the clinician suspects that NART performance may be impaired, Crawford, Allan, Cochrane & Parker (1990) built a regression equation to predict NART performance from the patient's background demographic information

1.5. The Wechsler Adult Intelligence Scale-Revised.

The WAIS-R (Wechsler, 1981) is the most commonly used test of intellectual ability in the English-speaking world. It is described by Lezak (1988b), as the principal, core measure of intellectual ability resorted to by clinical neuropsychologists, and the "workhorse of neuropsychological assessment". Ivnik, Malec, Smith, Tangalos, Peterson, Kokmen & Kurland (1992) refer to it as the "gold standard" in intellectual ability measurement. This position of pre-eminence is due to clinical preference for an individually administered composite test in battery format, the extensive standardisation sample, and it's elegant statistical properties. Implicit in the structure of the WAIS-R is a view of intellectual ability as a non-unitary, multi-dimensional entity. It is considered by Lezak (1995) to provide good information about overall intellectual functioning, and the presence or absence of significant intellectual impairment. It covers an extensive age range from 16-74 years and is comprised of six verbal subtests and five performance subtests, grouped on an *a-priori* basis. Not only are the standardisation samples large, they are highly representative of the US population. In addition, it's scaling and

reliability are generally excellent. Further, considerable emphasis is placed by some clinicians on discrepancies between test scores, and not only are the WAIS-R subtest scores on the same scale of measurement, they are standardised on the same population. This permits the potential analysis of discrepancies between subtest scores. Scores on other clinical measures can be transformed to *z*-scores or percentiles, as described by Anastasi, (1988, pp. 84-87), but the relevant standardisation samples must conform stringently so that variables which could affect test performance are controlled for.

1.5.1. Summary IQ scores, Subtest scores, and Factor scores.

Clinical interpretation of WAIS-R performance can be based on analyses at the level of the summary IQ scores, subtest scores or factor scores. There are divided opinions as to the utility of the summary IQ scores, with Lezak notable in her opinion that the discrepancy between Verbal and Performance IQ scores is unimportant. Conversely, Matarazzo and Herman (1985) have suggested that this discrepancy is the best-validated index of brain dysfunction. Lezak (1988a) suggests that IQ figures can obscure clinically important strengths and weaknesses. The fundamental problem as described by Lezak (1988a), is that patients with different pathologies will produce different patterns of strengths and weaknesses in terms of their profiles of subtest scores. According to Lezak's argument, it would be perfectly possible, for example to measure an individual's various garment sizes and compute an average garment size for that individual, which she would say is equivalent in absurdity to reducing the various facets of intellectual functioning to a single IQ concept. She quotes the example where patients

following closed head injury may be expected to perform badly on Digits Backwards, Arithmetic, and Digit Symbol. When these low scores are added in with their higher scores, to compute their IQ, their pattern of strengths and weaknesses is completely obscured. The profile, which is discarded, has both diagnostic value, and may be the basis for therapeutic intervention.

However, the interpretation of strengths and weaknesses in a subtest profile presents certain difficulties, since there has been little research on the interpretation of subtest scatter. Furthermore, subtest scores are more difficult to deal with statistically, and a commonly stated objection to replacing IQ scores with subtest profiles is that individual subtest scores are not as reliable as IQ scores (Matarazzo & Herman, 1985; Silverstein, 1981; Crawford, 1992). Lezak (1988a), commenting on the foregoing, states that it is possible to raise reliability coefficients by including more and more statistical variables into a global measure, but simultaneously rendering the global measure increasingly meaningless. She feels that the reduction in reliability is a price worth paying in return for the increased richness in clinical interpretation which attention to profiles or scatter affords.

Subtest scatter indices have been assessed in some detail by Crawford, Johnson, Michalkiw & Moore (1997) showing that in spite of significant intellectual deficit in a large brain injury sample relative to healthy controls, the *control* sample had a greater profile variability index (using the formula of McLean, Reynolds & Kaufman, 1990). The control sample exhibited *greater* subtest variability than the head injury subjects and casts doubt on the assumption that this is increased following concussional brain injury.

If these results are generalisable, intra-individual variability cannot be used as a basis for identifying acquired deficits, as in medico-legal assessment. This does not however undermine the general utility of evaluating a patient's strengths and weaknesses as a basis for intervention and rehabilitation.

A common error in clinical interpretation of WAIS-R data occurs where clinicians fail to appreciate the distinction between *reliable* and *abnormal* differences in subtest scores. Many clinicians in the past have used Table 13 from the WAIS-R manual (Wechsler, 1981) which provides critical values required for significance when comparing any subtest with any other. There are even computer programmes, which will compute all 55 possible combinations. This encourages the unwary clinician into *post hoc* inferences concerning *significant* (reliable) differences thrown up by the statistical method. To reduce the data to more manageable proportions on the one hand, and to preserve clinically significant attributes of the subtest profile on the other, Silverstein (1982) proposed a method for striking a balance between Type 1 and Type 2 error rates. Type 1 errors are inflated in situations in which multiple comparisons are made, and the WAIS-R manual (Wechsler, 1985) would have to be redrawn using a Bonferroni correction to control for the 55 comparisons involved for all possible subtest differences. This procedure would control for inflation of the Type 1 error rate, but would simultaneously reduce the power to detect subtest differences to an unacceptable level (i.e., the rate of Type 2 errors would be unacceptably high). Silverstein's ipsative method is that each subtest score is compared with that individual's *mean* subtest score. This reduces the statistical comparisons from 55 to only 11, and the application of the

Bonferroni correction for multiple comparisons satisfies the requirement to minimise both Type 1 and Type 2 errors in clinical inference. The ipsative method was also proposed independently by Knight & Godfrey (1984).

Silverstein (1984) presented tables of critical values for *abnormality* of subtest discrepancies.

The interpretation of abnormality of subtest scatter in clinical samples must be viewed against the base rates of scatter in normal samples, and to this end, Crawford & Allan (1996) have presented base-rate data on subtest scatter in a large healthy UK sample, following the publication of similar data in the US by Matarazzo, Daniel, Prifitera & Herman (1988). Given the finding of Crawford & Allan (1996) that 40% of the UK normal sample had subtest discrepancies at the 5% level of significance (i.e. they were *reliable* differences) and 25% had significant subtest differences at the 1% level, without this base-rate data, clinicians are in danger of committing Type 1 errors, that is, of inferring impairment from subtest profile discrepancies when, in fact, such variation occurs commonly in the healthy population. Crawford, Allan, McGeorge & Kelly (1997) have also presented tables to assess the *abnormality* of subtest differences for use with a number of WAIS-R short-forms.

1.5.2. Estimating premorbid subtest scores.

It follows on from the foregoing that clinical inference is enhanced by the ability to determine that an obtained subtest score is reliably and abnormally different from the mean of a subject's whole WAIS-R performance, or from the mean obtained by an

individual's performance on one of a number of recognised WAIS-R short-forms. The criterion validity of a number of these has recently been assessed by Crawford, Mychalkiw, Johnson & Moore (1996) in large healthy and clinical samples, and Crawford (1997) has published tables for assessing the statistical significance of subtest differences in WAIS-R short-forms. These include the four-subtest form proposed by Silverstein (1982), the five-subtest form proposed by Canavan, Dunn & MacMillan (1986) and the seven-subtest form proposed by Atkinson (1991). The last two of these are particularly notable in that they were derived to reflect the factor structure of the WAIS-R.

Base rate data on the abnormality of subtest scatter for WAIS-R short-forms have been presented by Crawford, Allan, McGeorge & Kelly (1997). However, in the clinical situation there may be a number of circumstances whereby a full-length WAIS-R is either impossible to administer or undesirable, and this might apply to recognised short-forms. Crawford, O'Carroll and Venneri (1998), for example, note that some of the WAIS-R test material is fundamentally unsuitable for subjects with reduced visual acuity (especially Picture Completion and Picture Arrangement). As Lezak (1988b; 1995) has noted, clinicians like herself resort to the WAIS-R as a starting point for generating hypotheses within a general clinical-theoretical ideological approach. Inevitably, the clinician will wish to collect clinical and psychometric data pertinent to a variety of neuropsychological functions including language, attention, conceptualisation and planning, memory, visual processing, and spatial reasoning, together with observations

of behaviour and social presentation. In a busy clinical service, the clinician may have to prioritise their investigation to fit into a finite time frame. Equally, subject fatigue may simply limit the amount of data that can be collected, lest the patient withdraw their cooperation. Neuropsychological data is gathered in a social context, and clinicians frequently encounter aspects of catastrophic reaction in formal clinical assessment. When the purpose of the investigation is to establish a clinical diagnosis, it may be appropriate on ethical grounds to terminate formal assessment unless further investigation leads to more effective management. If the purpose of the assessment is to conduct research, in most cases the clinician will be motivated to complete an assessment protocol.

There are then a number of circumstances in which the availability of data to support clinical inference from a single WAIS-R subtest scores would be of considerable utility, but always bearing in mind that subtests are, as noted, differentially impairment sensitive. If one wishes to analyse performance at the subtest level, this differential sensitivity of subtests presents practical difficulties. An alternative approach is to compare a patient's subtest score with an estimate of their premorbid performance on the same subtests. Furthermore, many clinicians may be interested in subtest scores as independent measures of particular cognitive domains in order to test particular clinical hypotheses. Until recently, a means of estimating premorbid WAIS-R subtest scores has not been available

McKinlay & Gray (1992), describing an approach to clinical inference for use in clinical interpretation and medico-legal reporting, advocate transforming a patient's scaled scores to age-graded scaled scores, converting these to individual IQ equivalents for comparison of the patient subtest by subtest with their peers (using an age-graded IQ equivalent of mean subtest score) and comparing the subtest IQ equivalents with NART-estimated VIQ and PIQ premorbid estimates. This approach can be criticised on at least two grounds. Firstly, subtest scatter is common in normal samples i.e. most normal subjects would have subtest scores deviating extremely from the mean for their age group; one cannot reasonably infer that observed deviations are necessarily due to acquired impairment. Secondly, subtests are not evenly related to NART estimated VIQ and PIQ, and are, as previously noted, differentially sensitive to impairment.

Recently, Paolo, Ryan, Troster & Hilmer (1996) have presented demographically based regression equations to estimate WAIS-R subtest scaled scores by combining the US WAIS-R standardisation data (Wechsler, 1981) and elderly WAIS-R standardisation sample (Ryan, Paolo & Brungardt, 1990). The data from 1880 and 130 persons respectively was entered into the analyses. The samples were stratified with regard to age, education, race, occupation, urban-rural domicile, and geographic region. Equal numbers of males and females were entered at each age level. The total normal sample was divided into two equal sized groups by random allocation. One group was used to develop preliminary regression equations and the second group for cross-validation. Stepwise multiple regression analyses were employed to determine which demographic

variables best predicted each of the 11 WAIS-R subtest scores. Following successful cross-validation of the initial equations, the two samples were combined, and new equations built.

The summary of results for the predicted and the actual scores for the equations derived from the combined normal sample of 2010 individuals is presented as Table 1.3 together with the results of comparing estimated and obtained scores, when the equations were applied in a clinical sample (N=247). All multiple Rs were significant and ranged from .47 for Digit Span to .69 for Digit Symbol. There were no statistically reliable differences between actual and estimated subtest scaled scores in the healthy sample. For all subtests, less than 50% of the variance was explained by the demographic variables. Their equations are also shown in Chapter 2.

The accuracy of the equations to predict scores within +/- 3 points for the sample was 76% for Arithmetic, 82% for Similarities and for Picture Completion, 83% for Digit Span and for Object Assembly, 85% for Comprehension, Picture Arrangement and for Digit Symbol, 86% for Block Design, 88% for Information and 89% for Vocabulary. Applying the equations to a general neurologically impaired sample of 247 patients, showed that for all subtests, estimated subtest score was significantly greater than obtained subtest score at the .001 level.

Table 1.3: Performance of Paolo *et al.* (1996) regression equations in the healthy sample (n=2010) and the clinical sample (N=247).

Subtest	Mult. R	a:% acc.	Healthy sample (N=2010) b:actual v estimated	Clinical sample (N=247) c:estimated>obtained
Information	.63	88	*	**
Digit Span	.47	83	*	**
Vocabulary	.65	89	*	**
Arithmetic	.58	76	*	**
Comprehension	.59	85	*	**
Similarities	.58	82	*	**
Pict. Comp.	.56	82	*	**
Pict. Arr.	.59	85	*	**
Block Design	.60	86	*	**
Object Ass.	.51	83	*	**
Digit Symbol	.69	85	*	**

Note a:% acc.= percentage of healthy sample predicted score accurate within +/- 3 scaled score points of actual score

b:*=for the healthy sample, all actual versus estimated scores non significant, with paired *t* test *p* values >.05.

c:**=for the clinical sample, all estimated scaled scores were significantly greater than obtained scaled scores $p<.001$.

The key issue thereafter is the base rates of both *reliable* and *abnormal* differences in the normal sample but this issue was not specifically discussed by the authors. Although significantly more brain-damaged individuals' obtained scaled scores fell below the 90% and 95% confidence limits of the estimated scaled score for all subtests, 35% of the normal sample displayed possible decline on one or more subtest at the 90% confidence limit and 16% at the 95% confidence limit. Conversely, 78% of the clinical sample displayed possible decline on at least one subtest at the 90% confidence limit, and 60% did so at the 95% confidence limit. Notwithstanding the overlap between the two groups, these are impressive statistics when one considers that less than 10% of the brain-damaged group exhibited a Verbal-Performance discrepancy of more than 21 points (Ryan, Paolo & Van Fleet, 1994). Paolo *et al.* note that consistent with other research employing demographic equations to predict premorbid ability, their equations tend to underestimate ability at the higher extreme of ability and to overestimate it at the other extreme. They say that in practice this would be manifest by inferring relatively low scaled scores (i.e. 6 or less) as evidence of deterioration whilst tending to regard scaled scores of 14 or more as normal. They suggest that the clinician may need to resort to other corroborating information to determine whether or not an obtained score may represent deterioration from a previously higher level of functioning. In the present study, the possible contribution of NART data to this problem will be examined (Chapter 2).

1.5.3. WAIS-R Factor scores.

A number of factor analytic studies using the Wechsler scales have been reviewed by Leckliter, Matarazzo & Silverstein (1986). Factor analyses of WAIS-R data have consistently suggested three underlying ability dimensions: a *verbal* factor (V), on which Information, Vocabulary, Comprehension and Similarities have high loadings, a *perceptual organisation* factor (PO), on which Block Design and Object Assembly have the principal loadings, and a third factor, termed *attention-concentration* or *freedom from distractibility* factor (A/C), consisting of loadings from Arithmetic, Digit Span and often, Digit Symbol (Crawford, Allan, Stephen, Parker & Besson, 1989; Leckliter, Matarazzo & Silverstein, 1986). Conventional IQ summary scores clearly do not have optimal construct validity in view of the previous finding, and this is not surprising given that the subtests were allocated to Verbal and Performance sub-scales on purely intuitive grounds. The models derived from factor analyses should in principle produce better measures of underlying verbal and non-verbal aspects of cognition than conventional summary IQ measures. Review of the WAIS and laterality of lesion by Bornstein & Matarazzo (1982) has shown that VIQ is typically lower than PIQ in patients with well-defined left brain lesions and the converse relationship is found with right brain lesions. Bornstein (1983) demonstrated the same relationships with WAIS-R. However, there are so many exceptions to the general group relationships, that the findings are unusable in the individual case. Using factor IQ scores, may provide a better measure of laterality of

change in function in patients with unilateral hemisphere brain impairment, and evidence in support of this comes from a study by Lawson, Inglis & Stroud (1983).

It has also been established that the factor structure is extremely robust in normal samples in different cultures and, importantly, also in clinical samples (e.g., Atkinson, Cyr, Doxey & Vigna, 1989). This demonstrates that the underlying ability dimensions are robust, in that when intellectual ability changes as a result of injury or disease, it does so with respect to those underlying ability dimensions. The attraction of factorially derived composite measures is in their status as broad indicators of current intellectual functioning against which other more specific neuropsychological measures can be compared (Crawford, Johnston, Mychalkiw & Moore, 1997). Although the reliabilities of factor scores are marginally less than those of WAIS-R summary IQ scores, they are better suited to this role as broad indicators of intellectual functioning in clinical samples via their superior construct validity.

There are two approaches to the calculation of factor scores. The first method is simply to derive a WAIS-R short form, which reflects the factor loadings present in the factor analyses of standardisation data. For example, for a three factor solution, the WAIS-R would be broken down into a Verbal Factor short form consisting of Vocabulary, Information, Comprehension, and Similarities; a Performance Factor short form consisting of Block Design and Object Assembly; and an Attention-Concentration / Freedom from Distractibility Factor short form comprising Digit Span and Arithmetic. The second approach is to apply weights to each of the subtests according to their loading on each of the factors.

Until recently, there has been no convenient method for factor scoring the WAIS-R, as well as the absence of data for interpreting the reliability and abnormality of discrepancies between factors. Atkinson (1991) has now facilitated the use of factor scores in clinical practice based on work using the WAIS-R US standardisation sample. He provided equations for their computation, tables of standard deviations, reliability coefficients and standard errors of estimation and prediction. He has also provided tables of critical values for the assessment of significance and abnormality of differences amongst factor scores, specifying their confidence intervals, and outlining a method for determining the significance and abnormality of deviation quotients.

Crawford, Johnston, Michalkiw & Moore (1997) have presented data on the relative utility of IQ scores, subtest scatter indices and factor scores (weighted method) in discriminating between normal and impaired subjects. Factor scores were significantly superior to both IQ scores and scatter indices, with the latter performing only at chance levels.

Given the arguments in favour of factor scores as summary indices of intellectual ability it is surprising that no research has been done to provide a means of estimating premorbid factor scores. In the present study models for the prediction of factor scores from demographic information and NART data alone and in combination will be evaluated, extending the work of Atkinson (1991) and removing the last barrier to their adoption in clinical practice.

1.6. Study aims and hypotheses.

1. The preliminary aim of the present study is to replicate the study by Paolo *et al.* (1996) in a large, healthy, stratified UK sample to evaluate the criterion validity of demographic regression models to explain variance in individual WAIS-R subtest scores. This will permit a comparison of the relative utility of demographic regression models in the UK and US.

2. Combining NART and demographics improves prediction of WAIS-R summary IQ scores in UK samples, and the second aim of this study is to establish if combining NART and demographic variables significantly improves prediction of subtest performance in the normal sample.

3. The third aim is to extend the second part of the Paolo *et al.* (1996) study, to examine the utility of estimating premorbid WAIS-R subtest scores in a UK clinical sample, using regression models combining NART and demographic variables. It is hypothesised that using estimated premorbid scores in conjunction with obtained scores will improve discrimination over that achieved by obtained scores alone.

4. As an alternative to WAIS-R summary IQs, the fourth aim is to evaluate the success of models incorporating demographic information and models combining NART and

demographic information in explaining variance in WAIS-R factor scores in a stratified, healthy UK sample.

5. Finally, it is proposed to examine the utility of the regression models developed above in a UK clinical sample. It is hypothesised that the combination of obtained factor scores and estimated premorbid factor scores will be superior in terms of discrimination between healthy and impaired subjects than obtained factor scores alone and, further, that the combination of obtained and estimated premorbid factor scores will be superior to the use of obtained and estimated premorbid summary IQ scores.

CHAPTER 2

ESTIMATING PREMORBID WAIS-R SUBTEST SCORES

2.1. Introduction.

As discussed previously, there may be a number of circumstances when a clinician wishes to make inferences from only one or a small number of individual WAIS-R subtest scores. For example, McKinlay & Gray (1992) advocate converting Age Scaled Scores to individual IQ equivalents, to facilitate a comparison between that individual, subtest by subtest, with NART estimated *summary* premorbid IQ. This procedure ignores the base rate of subtest scatter in the healthy population, where widely dispersed IQ equivalents would be obtained by the majority of *healthy* subjects. This approach to clinical interpretation would tend to inflate the apparent detection of impairment in clinical subjects. Furthermore, NART will provide a variably accurate premorbid estimate, and therefore the size of discrepancy required to be abnormal would be expected to vary from subtest to subtest. In failing to take account of the variable relationships between subtest scores and NART scores, the approach advocated by McKinlay & Gray (1992) is pseudo-quantitative, and inevitably, likely to be inaccurate.

This chapter presents data attempting to replicate the study by Paolo, Ryan, Troster & Hilmer (1996), and evaluates the utility of demographic predictor models in estimating premorbid WAIS-R subtest scaled scores based on a stratified UK healthy sample. The demographic models will be compared with models combining demographic variables and NART.

Paolo *et al.* (1996) presented demographically-based regression equations to predict subtest scaled scores combining the WAIS-R standardisation data (N=1880) and the (N=130) old age standardisation sample (Ryan, Paolo, & Brungardt, 1990). The combined sample consisted of 2010 individuals from age 16 to 96 stratified on the basis of age, education, race, occupation, urban-rural residence, and geographic region. Demographic variables were selected on the basis of separate, stepwise multiple regression analyses. The equations generated from this procedure are shown as Table 2.1, together with their associated multiple Rs. There were no statistically significant differences between actual and estimated subtest scaled scores (all paired *t* test *p* values > .05; see also Table 1.3).

The aim of the present study is to evaluate the criterion validity of demographically based predictor models in estimating premorbid performance on individual WAIS-R subtests for use in the UK population, to subsequently include data from a present ability measure (NART), and to report it's effect on predictive accuracy. It is hypothesised that NART will significantly improve prediction, particularly for verbal subtests.

Table 2.1: Paolo *et al.* equations to predict WAIS-R subtest scores with multiple Rs.

Subtest	Equation	R.
Information =	age(0.131)+educ(1.184)+sex(0.854)+race(1.084) +job(0.247)+1.174.	.63
Digit Span =	age(-0.092)+educ(0.795)+race(0.744)+job(0.235)+5.36.	.47
Vocabulary =	age(0.184)+educ(1.23)+race(1.341)+job(0.271)+1.669.	.65
Arithmetic =	educ(0.981)+sex(0.945)+race(1.448)+job(0.254)+2.09.	.58
Comprehension =	age(0.107)+educ(1.125)+race(1.344)+job(0.27)+2.568.	.59
Similarities =	age(-0.138)+educ(1.156)+race(1.163)+job(0.21)+3.531.	.58
Picture Compl. =	age(-0.388)+educ(0.84)+sex(0.508)+race(1.468)+5.531.	.56
Picture Arr.=	age(-0.409)+educ(0.757)+sex(0.385)+race(0.755) +job(0.148)+region(0.139)+5.733.	.59
Block Design =	age(-0.355)+educ(0.71)+sex(0.655)+race(1.765) +job(0.21)+region(0.185)+4.127.	.60
Object Ass. =	age(-0.382)+educ(0.641)+race(1.461)+6.923.	.51
Digit Symbol =	age(-0.599)+educ(0.856)+sex(-0.947)+race(0.877) +job(0.251)+7.971.	.69

Note: Demographic codes were adapted from the WAIS-R manual (Wechsler, 1981). There were 12 bands for age, six for years of full-time education, two for gender and racial grouping, six for job, four for region of domicile and two for residence (rural or urban).

Intuitively, there may be expected to be relationships amongst independent variables. For example, it seems likely that an individual of a given IQ, raised in a verbally sophisticated environment will score higher on word reading tasks compared with another individual of the same IQ, but raised in a verbally impoverished one. Similarly, there may be a relationship between age and education, in view of the steadily increasing proportion of young people staying on in higher education compared to their parents. To test for the presence of significant interactions, each independent variable will be paired in all possible combinations, and multiplied to form new variables according to the procedure described by Cohen & Cohen (1982). The potential contribution of these new variables to predictive accuracy will be evaluated in further regression analyses following forced entry of the primary demographic predictor variables. The present writer is not aware of any published work in which the potential statistical impact of this issue has been evaluated.

2.2 Method.

Professor J R Crawford, University of Aberdeen, provided a database concerning the cognitive performance and demographic characteristics of 451 healthy individuals from which the healthy sample employed in the present study was selected. Not all subjects in the database had completed all measures of interest. Subjects were selected for inclusion in the study sample if they had completed a full WAIS-R and a NART administered by a team of research assistants. They were ascertained to be free from developmental or acquired neurological, psychiatric or sensory disability by self report

and by superficial visual inspection at interview and had been recruited from a variety of sources including local and national public and private companies, clubs and associations (e.g., old-age pensioner's clubs, fishing clubs, attenders of community centres etc.). Most subjects were Aberdeen urban dwellers and received a small honorarium in return for their participation. A sub-sample of eligible cases was formed to be representative of the adult UK population, and consisted of 245 individuals (122 males and 123 females).

Mean age of the sample was 43.09 (SD = 17.97), with a range from 16 to 83 years. The social class of each subject was recorded, being derived from their present or previous occupation, provided this reflected the highest level of occupational activity for that individual during their working life (to control for the potential effect of those winding down to retirement), using the Classification of Occupations (OPCS, 1980). Married women were coded by their husband's occupation. The contemporary unemployed were coded by their highest previous occupation. The mean years of education for the sample was 12.7 (SD = 3.0), with a range from 7 to 21. In addition to years of full-time education, subjects were credited with 0.25 of a year for every year attendance at day or evening classes, provided these led to a recognised qualification.

The sample was formed to be broadly representative of the UK adult population, with respect to the distributions of social class, age, and sex. To examine the extent to which this was achieved, the following comparisons were made between the sample and the UK adult population:

Firstly, the social class distribution of the present sample was compared with that of the UK adult population according to the 1981 census. The social class distributions of the present sample and those pertaining to the 1981 census are shown as Table 2.2. A Chi-square goodness-of-fit test showed that the social class distribution in the present sample did not differ significantly from the population distribution ($\chi^2 = 5.163$, d.f. = 4, $p = .271$).

Secondly, a similar procedure was adopted to examine the representativeness of the sample in terms of age distribution. Nine age bands were formed, corresponding to those adopted for the WAIS-R standardisation sample, except that the 70-74 age band was replaced with a 70+ band. This data is shown as Table 2.3. A Chi-square test revealed that the sample and expected distributions did not differ significantly ($\chi^2 = 12.677$, d.f. = 8, $p = .0123$).

Finally, a Chi-square test showed that the sex distribution did not differ from that derived from the census ($\chi^2 = 0.062$, d.f. = 1, $p = .803$).

Table 2.2. Social class distributions in the present sample and UK population.

	Social class				
	1	2	3	4	5
General adult UK population	5	23	48	18	6
Present sample	17	60	107	41	20

Table 2.3: Distributions of age in the UK population and in the present sample.

	Age distribution								
	16-17	18-19	20-24	25-34	35-44	45-54	55-64	65-69	70+
This study	6	7	27	55	49	33	24	16	28
UK census	4.5	4.3	9.7	18.7	15.8	14.8	14.61	6.5	11.3

2.3 Procedure.

All subjects had been administered a full-length WAIS-R according to standard procedures (Wechsler, 1981; Lea, 1986), and had completed the NART (Nelson, 1982). In order to achieve comparability with the Paolo *et al.* (1996) study, all analyses in the present study were performed on WAIS-R scaled scores rather than age-graded scores. The following procedures were adopted to test hypotheses:

(1) Can demographic variables account for a significant proportion of the variance in WAIS-R subtest scores?

To evaluate the ability of demographic variables to estimate variance in WAIS-R subtest scores, the independent demographic variables age, sex, class, and education were entered simultaneously in separate regression analyses for each WAIS-R subtest.

(2) Does combining NART and demographic information significantly improve prediction of WAIS-R subtest scores?

In order to determine if the combined use of NART and demographic information significantly improves prediction over that achieved by demographic information alone, a second series of analyses were conducted, where NART scores were entered following forced entry of the demographic variables in separate hierarchical two-stage multiple regression analyses. Thus it was possible to examine the additional contribution of the NART variable to predictive accuracy in respect of each subtest. Further two-stage multiple regression analyses were conducted where NART scores were entered first, followed by forced entry of the demographic variables, to establish if the practical effort involved in the collection of demographic information and calculation of the models could be justified statistically.

(3) Interpreting estimated-obtained discrepancies.

Knowledge of the base rates of discrepancies between estimated and obtained test scores in the healthy population is indispensible to clinical interpretation. To assist in the interpretation of discrepancies between estimated and obtained subtest scores critical values were calculated for four levels of significance by multiplying the corresponding *z* value (one tailed) by the standard error of estimate (SE_{est}) for each of the relevant regression equations.

(4) Do statistical interactions between predictor variables significantly improve predictive accuracy?

In order to determine if the modelling of potential interactions between predictor variables significantly improves prediction over that achieved using the Demographics and NART plus demographics models the contributions of all possible paired two-way interactions was calculated by further regression analyses. According to the method described by Cohen & Cohen (1982), a series of new variables were created by multiplying each of the predictor variables with the other predictor variables. Thus, in the case of the NART plus demographics models, there were five predictor variables and thus ten additional variables were created to represent all possible two-way interactions. These variables were entered in separate two stage hierarchical regression analyses following forced entry of the NART and primary demographic predictor variables. In the case of the models using demographic variables alone, the same procedure was adopted but with exclusion of NART as a primary predictor variable and exclusion of variables expressing NART and demographic interactions. In view of the potentially large number of calculations involved, it was decided to restrict the analyses to two-way interactions unless these demonstrated an important effect.

2.4 Results

Summary statistics for the demographic predictor variables and the psychometric test performance of the sample are shown in Table 2.4.

Table 2.4: Demographic characteristics and psychometric performance of the healthy sample (N=245).

	Mean	SD	Range
Age	43.09	17.97	16-83
Education	12.71	3.00	7-21
NART errors	19.50	9.49	3-41
FSIQ	103.07	13.26	71-140
VIQ	103.02	12.85	73-133
PIQ	102.42	13.36	67-139
Information	9.79	3.08	3-18
Digit Span	10.83	2.77	3-19
Vocabulary	10.05	2.52	4-18
Arithmetic	10.97	3.02	5-17
Comprehension	9.81	2.76	2-18
Similarities	9.00	2.69	2-15
Picture Completion	8.94	2.87	3-17
Picture Arrangement	9.51	3.19	2-17
Block Design	9.63	3.03	4-19
Object Assembly	8.44	2.79	1-18
Digit Symbol	8.93	2.86	2-18

2.4.1. Estimating premorbid WAIS-R subtest scores.

Table 2.5 presents the proportion of variance (R^2) explained by the predictor variables for each WAIS-R subtest and the corresponding standard errors of estimate (SE_{est}) for the four prediction models: US demographic, UK demographic, UK NART plus demographics combined, and NART. The Paolo *et al.,* (1996) US data is included to permit direct comparison with the UK demographic models. The UK NART plus demographics models were created in two-stage regression analyses with NART added to demographics and with demographics added to NART. The SE_{est} statistic is a measure of variability about the regression line, and as such is an index of the error in prediction. It is analogous to standard deviation (SD), and the size of SE_{est} relative to the SD of y is a function of the strength of the correlation (r), between the predictor variable (x) and y, according to the following equation:

$$SE_{est} = SD_y \sqrt{(1 - r^2)}$$

It can be seen from Table 2.5 that the UK demographic models predict from 20 to almost 50% of the variance in subtest performance. For all but two of the subtests, the UK demographic models predict more, or an equivalent amount, of the variance in subtest scores than the equivalent Paolo *et al.* (1996) US demographic models.

Table 2.5: Summary statistics for estimated subtest scores for demographic models, combined models, and the NART models.

	US DEM.		UK DEM.		NART + DEM.		NART	
Subtest	R^2	SE_{est}	R^2	SE_{est}	R^2	SE_{est}	R^2	SE_{est}
Inf.	.397	2.31	.486	2.23	.649	1.84	.558	2.05
D.Span	.221	2.54	.201	2.50	.294	2.35	.248	2.41
Voc.	.423	2.28	.474	1.84	.706	1.38	.681	1.43
Arith.	.376	2.44	.393	2.37	.452	2.26	.301	2.53
Comp.	.348	2.48	.375	2.20	.463	2.05	.396	2.15
Sim.	.336	2.61	.346	2.19	.482	1.95	.392	2.10
P.C.	.314	2.66	.240	2.52	.320	2.39	.236	2.51
P.A.	.325	2.60	.299	2.70	.356	2.59	.218	2.83
B.D.	.360	2.42	.449	2.27	.466	2.24	.159	2.79
O.A.	.260	2.47	.262	2.42	.278	2.40	.106	2.64
D.Sym	.476	2.42	.496	2.05	.520	2.00	.068	2.77

Note: The NART plus demographics models were created in two-stage regression analyses by adding NART to demographics and by adding demographics to NART; All changes in F following the addition of NART to demographics were significant at $p<.001$, except Block Design (F for R^2 change $p=.008$), Object Assembly (F for R^2 change $p=.02$), and Digit Symbol (F for R^2 change $p=.001$). All changes in R^2 following the addition of demographics to NART were significant at $p<.001$, except Digit Span (F for R^2 change $p=.004$) and Vocabulary (F for R^2 change $p=.001$).

The addition of NART to the regression models results in generally enhanced predictive ability with changes in R^2 significant at $p<.001$ level, except for Block Design ($p=.008$), Object Assembly ($p=.02$) and Digit Symbol ($p=.001$). In all cases, the addition of NART to demographics improved prediction and this was an aim of the present study. The largest changes in the variance explained occurred in Vocabulary (24%) and in Information (16%). There were more modest improvements in relation to Similarities and Comprehension (13% and 8%). Improvements in relation to the Performance subtests were uniformly modest ranging from 2% for Block Design, Object Assembly and Digit Symbol, to 6% for Picture Arrangement and 8% for Picture Completion. Where demographics were added after NART in the two-stage regression analyses, the changes in R^2 were significant at $p<.001$ except for Digit Span ($p=.004$) and for Vocabulary ($p=.001$).

In general one can conclude that where it is possible to collect the information, the NART plus demographics models provide the most accurate estimates of subtest performance. The NART plus demographics models outperform the other models with respect to predicting variance in all WAIS-R subtests.

The NART models outperform the UK demographics models for all Verbal subtests except Arithmetic. The NART model estimates only 7% of the variance in Digit Symbol performance, and within the WAIS-R Performance subtests, the highest estimating model is that for Picture Completion which accounts for a modest 24% of the variance.

2.4.2. UK Demographics models.

The results of the regression analyses permitted the construction of UK demographic equations for predicting premorbid WAIS-R subtest scores. These equations are presented in Table 2.6. The UK equations for predicting premorbid subtest scores from NART and demographics combined are shown as Table 2.7.

Table 2.6: UK demographic equations to predict premorbid WAIS-R subtest scores.

Information = 3.999+(0.037×age)-(1.347×sex)+(0.573×educ)-(0.367×class).
Digit Span = 9.832-(0.014×age)+(0.108×sex)+(0.245×educ)-(0.570×class).
Vocabulary = 4.027+(0.032×age)-(0.104×sex)+(0.489×educ)-(0.481×class).
Arithmetic = 10.833-(0.014×age)-(1.057×sex)+(0.345×educ)-(0.693×class).
Comprehension = 7.700+(0.012×age)-(0.160×sex)+(0.346×educ)-(0.860×class).
Similarities = 8.599-(0.016×age)-(0.313×sex)+(0.289×educ)-(0.717×class).
Picture Completion = 8.033-(.0017×age)-(.611×sex)+(.291×educ)-(.384×class).
Picture Arrangement = 9.330-(0.039×age)-(0.120×sex)+(0.308×educ)-(0.634×class).
Block Design = 11.810-(0.056×age)-(0.891×sex)+(0.271×educ)-(0.629×class).
Object Assembly = 10.140-(0.041×age)-(0.570×sex)+(0.182×educ)-(0.476×class).
Digit Symbol = 12.313-(0.101×age)+(0.880×sex)+(0.102×educ)-(0.562×class).

2.4.3. UK NART plus demographics models.

Table 2.7: Equations for predicting premorbid WAIS-R subtest scores from NART and demographics combined.

Information = 11.847+(0.01618×age)-(1.455×sex)+(0.232×educ)-(0.03150×class)-(0.185×NART).
Digit Span = 15.170-(0.028×age)+(0.034×sex)+(0.012×educ)-(0.299×class)-(0.126×NART).
Vocabulary = 11.664+(0.012×age)-(0.210×sex)+(0.156×educ)-(0.093×class)-(0.180×NART).
Arithmetic = 15.479-(0.027×age)-(1.122×sex)+(0.143×educ)-(0.457×class)-(0.110×NART).
Comprehension = 12.838-(0.0023×age)-(0.231×sex)+(0.122×educ)-(0.599×class)-(0.121×NART).
Similarities = 14.841-(0.033×age)-(0.399×sex)+(0.018×educ)-(0.400×class)-(0.147×NART).
Picture Completion = 13.128-(0.031×age)-(0.681×sex)+(0.069×educ)-(0.125×class)-(0.120×NART).
Picture Arrangement = 14.133-(0.05216×age)-(0.186×sex)+(0.0981×educ)-(0.390×class)-(0.113×NART).
Block Design = 14.238-(.0063×age)-(.925×sex)+(.166×educ)-(.506×class)-(.0057×NART).
Object Assembly = 12.393-(0.047×age)-(0.601×sex)+(0.084×educ)-(0.361×class)-(0.053×NART).
Digit Symbol = 15.113-(0.108×age)+(0.841×sex)-(0.019×educ)-(0.420×class)-(0.066×NART).

2.4 4. UK NART models.

The equations for estimating premorbid WAIS-R subtest scores from NART score are presented in Table 2.8.

Table 2.8: UK Equations for predicting WAIS-R subtest scores from NART

Information = 14.516 - (0.242 x NART).
Digit Span = 13.663 - (0.145 x NART).
Vocabulary = 14.323 - (0.219 x NART).
Arithmetic = 14.370 - (0.175 x NART).
Comprehension = 13.383 - (0.183 x NART).
Similarities = 12.458 - (0.177 x NART).
Picture Completion = 11.799 - (0.147 x NART).
Picture Arrangement = 12.569 - (0.157 x NART).
Block Design = 12.118 - (0.127 x NART).
Object Assembly = 10.309 - (0.09582 x NART).
Digit Symbol = 10.453 - (0.07827 x NART.

2.4.5. Estimated-Obtained discrepancies.

To assist in the clinical interpretation of discrepancies between predicted and observed subtest scores, the SE_{est} for each subtest was multiplied by the value of *z* (one tailed) corresponding to the 15%, 10%, 5%, and1% levels of statistical significance which a discrepancy must exceed to be regarded as unusual relative to that particular size of discrepancy occurring in the general (healthy) population. One tailed values are appropriate since in measuring impairment, one is only interested in obtained score lower than estimated. These critical values for the UK Demographics subtest equations are shown as Table 2.9; those pertaining to the UK NART plus demographics combined equations are shown in Table 2.10; and those pertaining to the UK NART equations are shown as Table 2.11.

Table 2.9: Size of discrepancy between obtained and estimated subtest score required for four levels of significance; UK Demographics models.

	15%	10%	5%	1%
Information	2.30	2.85	3.66	5.17
Digit Span	2.58	3.20	4.10	5.80
Vocabulary	1.86	2.36	3.02	4.23
Arithmetic	2.44	3.03	3.89	5.50
Comprehension	2.27	2.82	3.61	5.10
Similarities	2.16	2.80	3.59	5.08
Picture Completion	2.60	3.21	4.13	5.85
Picture Arrangement	2.78	3.46	4.43	6.26
Block Design	2.34	2.91	3.72	5.27
Object Assembly	2.49	3.10	3.97	5.61
Digit Symbol	2.11	2.62	3.36	4.76

Note: Values are one-tailed.

Table 2.9 is designed to be used to aid interpretation of WAIS-R data from individual subtests where the use of the NART is either not possible or is inappropriate.

Table 2.10: Size of discrepancy between obtained and estimated subtest score required for four levels of significance; NART plus demographics model.

	15%	10%	5%	1%
Information	1.90	2.36	3.02	4.27
Digit Span	2.42	3.01	3.85	5.45
Vocabulary	1.42	1.77	2.26	3.20
Arithmetic	2.33	2.89	3.71	5.24
Comprehension	2.11	2.62	3.36	4.76
Similarities	2.01	2.50	3.20	4.52
Picture Completion	2.46	3.06	3.92	5.54
Picture Arrangement	2.67	3.32	4.25	6.01
Block Design	2.31	2.87	3.67	5.20
Object Assembly	2.47	3.07	3.94	5.57
Digit Symbol	2.06	2.56	3.28	4.64

Note: All values are one-tailed.

Table 2.10 is designed to be used in the interpretation of WAIS-R data in conjunction with the NART and demographics. As we have seen, the combination of NART and demographic information aids predictive accuracy with regard to premorbid subtest scores, and is therefore preferred over demographics alone provided that the particular clinical circumstances permit.

To illustrate, take the example of a patient who obtains a score of 11 on the Vocabulary subtest. The patient is female (coded 2), aged 28 years, is employed as a Solicitor (coded as social class 1), has 18 years of education, and a NART error score of 4. Entering these demographic and NART data into the combined equation for the Vocabulary subtest from Table 2.7, yields an estimated premorbid subtest score of 13.575. The discrepancy between the estimated and obtained score (13.575 - 11) is 2.575. Referring to Table 2.9, it can be seen that the discrepancy exceeds the critical value for significance at the 5% level of 2.26, but is less than the critical value for significance at the 1% level of 3.20 (i.e. 1% of the normal population would be expected to produce a discrepancy of this magnitude).

Occasionally demographic information may be unavailable or unusable, and Table 2.11 shows critical discrepancies required for significance when estimating from NART data alone. Because of the weaker power of the NART models the magnitude of discrepancies required for significance is generally larger than those required when estimating from the combination of NART and demographic information with reference to Table 2.10.

2.4.6 Interactions between predictor variables.

As noted, there may be expected to be relationships amongst independent variables that explain variance in the criterion variables (i.e. the subtests). For example, it seems likely that an individual of a given IQ, raised in a verbally sophisticated environment will score higher on word reading tasks compared with another individual of the same IQ, but raised in a verbally impoverished one.

Table 2.11: Size of discrepancy between estimated and obtained subtest score for four levels of significance; NART model.

Subtest	15%	10%	5%	1%
Inf.	2.11	2.62	3.36	4.76
D.Span.	2.48	3.08	3.95	5.59
Vocab.	1.47	1.83	2.35	3.32
Arith.	2.61	3.24	4.15	5.87
Comp.	2.21	2.75	3.53	4.50
Sim.	2.16	2.69	3.44	4.87
P.C.	2.59	3.21	4.12	5.82
P.A.	2.91	3.62	4.64	6.57
B.D.	2.87	3.57	4.58	6.47
O.A.	2.72	3.38	4.33	6.12
D.Sym.	2.85	3.55	4.54	6.43

Note: these values are one-tailed.

Similarly, there may be a relationship between age and education, in view of the steadily increasing proportion of young people staying on in higher education compared to their parents. Separate analyses were conducted in relation to the Demographics models, and the NART plus demographics models. For the Demographics models, the primary model

R^2, the R^2 change% following incorporation of the relevant interaction variables, and significance of R^2 change are presented in Table 2.12. Corresponding data for the NART plus demographics models are shown as table 2.13.

Table 2.12: Effect of interactions* on variance: Demographics models.

Subtest	R^{2A}	R^{2B}	R^2change%	*p* change in F
Inf.	.486	.498	1.2	.460
D.Span	.201	.211	1.0	.800
Voc.	.474	.512	3.8	.006
Arith.	.393	.414	2.2	.202
Comp.	.375	.423	4.8	.004
Sim.	.346	.381	3.5	.042
P.C.	.240	.281	4.1	.042
P.A.	.299	.318	1.9	.378
B.D.	.449	.460	1.0	.620
O.A.	.262	.310	4.8	.014
D.Sym.	.496	.521	2.5	.062

Note: R^{2A}= variance estimated by Demographics models; R^{2B}= variance estimated following entry of interaction variables*; interaction variables*= education x class, age x sex, age x education, education x sex, age x class, sex x class.

Table 2.13: Effect of interactions* on variance; NART plus demographics models.

Subtest	R^{2A}	R^{2B}	R^2 change %	Sig. of change in F.
Inf.	.649	.666	1.7	.315
D.Sp.	.294	.317	2.2	.678
Voc.	.706	.722	1.6	.220
Arith.	.452	.470	1.8	.649
Comp.	.463	.504	4.1	.049
Sim.	.482	.508	2.5	.307
P.C.	.320	.353	3.3	.324
P.A.	.356	.380	2.4	.545
B.D.	.466	.483	1.7	.658
O.A.	.278	.328	4.9	.086
D.Sym.	.520	.547	2.7	.198

. Note: R^{2A}= estimated variance NART plus demographics models; R^{2B}= estimated variance following entry of interactions*; Interaction variables*= age x NART, education x class, sex x NART, age x sex, education x NART, age x education, sex x class, age x class, NART x class, education x sex.

In respect of the effect of interaction on the Demographics models shown in Table 2.12, their incorporation has a generally modest positive effect on the estimation of variance which reaches significance with regard to Comprehension and Vocabulary. With regard to the effect of this procedure within the NART plus demographics models,

it can be seen from Table 2.13 that none of the changes in variance explained reaches statistical significance.

2.5. Discussion.

The preliminary aim of the present study was to replicate the work of Paolo *et al.* and to evaluate the criterion validity of models incorporating demographic variables i.e. to assess their ability to explain variance in WAIS-R subtest scores in a large, stratified UK healthy sample.

This aim was achieved using a series of statistical regression analyses with generally favourable results, with the UK demographic model performing as well as, or better than, the US models reported by Paolo *et al.,* (1996).

The procedure adopted in the present study was designed to permit a number of further analyses to establish if the incorporation of NART would significantly improve prediction over demographically based models alone, and in which circumstances. Bearing in mind the work of Crawford & Allan (1996) which showed that 40% of a healthy sample had *reliable* WAIS-R subtest discrepancies at the 5% level of significance, and that 25% had significant subtest discrepancies at the 1% level, the practice of relating individual scaled score subtest IQ equivalents (e.g. McKinlay & Gray, 1992) to global premorbid estimates of premorbid ability (e.g. NART estimated VIQ and PIQ), cannot be justified. Specific equations are presented here combining NART and demographics for the estimation of individual subtests, and it is recommended that these are resorted to in preference.

For those circumstances when the use of a present ability measure is inappropriate, demographic equations are presented for use in UK clinical practice, replicating the work in the US by Paolo, Ryan Troster & Hilmer (1996).

It must be noted with reference to Table 2.5, that using NART alone was a better predictor of subtest score than demographic information for all Verbal subtests apart for Arithmetic in a healthy UK sample. The NART was examined alone and in combination with the other measures to illustrate the independent predictive ability of word reading as well as in conjunction with demographic variables. The attraction of the NART as noted by O'Carroll (1995), is that it can provide a quick and easy estimate of premorbid WAIS-R intelligence, and as demonstrated in this study, is more accurate than demographic information on it's own in some but not all circumstances.

As we have seen, with respect to UK samples, the NART plus demographics models are most successful in estimating variance in premorbid WAIS-R subtest scores. This was the second aim of the present study.

Relative to the US demographic equations, the UK NART plus demographics models have an impressively enhanced ability to predict variance in particular subtests, and in no cases was prediction worse. Prediction of variance in Vocabulary rises by 29%, Information by 15%, Similarities by 14%, Comprehension and Block Design by 11%, Digit Span by 8%, Arithmetic by 7%, Picture Arrangement and Object Assembly by 3%, and lastly, Picture Completion by only 1%. Prediction of individual subtest variance, is in some cases substantially better than has been achieved by researchers in the US predicting WAIS-R summary IQ measures from demographic information alone.

To assist in the clinical interpretation of discrepancies between estimated premorbid and observed subtest scores, the present study complements the work on the base rates of subtest discepancies presented by Crawford & Allan (1996). Specifically, a range of tables are presented showing the critical magnitude of discrepancy required for significance relative to that found in the healthy population. Tables are presented for use with demographic information, for NART and demographics combined, and for NART alone, to cover the range of circumstances encountered in clinical practice. As noted, Crawford & Allan (1996) demonstrated that 40% and 25% of a stratified UK healthy sample had subtest discrepancies at the 5% and 1% levels of significance respectively, drawing attention to the danger of over-inference when encountering such discrepancies in clinical subjects. The work presented in this chapter also provides an alternative to simple clinical judgement in making inferences from apparent impairment in test performance by formalising background information in models in which the statistical effect of demographic information is defined objectively.

Following the method of Cohen & Cohen (1992), incorporating interactions between pairs of predictor variables into the regression models led to a generally modest improvement in the prediction of variance in WAIS-R subtest performance which, with few exceptions, did not reach statistical significance. Intuitively, it would seem likely that an individual raised in a verbally rich social environment would be expected to have superior word-reading ability than another individual, of equivalent IQ, raised in a relatively verbally impoverished social environment. Similarly, there may be expected to exist interactions between educational opportunities and social class, tending to weaken

the relationship between IQ and years of education because of the intervention of a purely social variable. In spite of this, the generally modest improvements in prediction achieved using interactions between predictor variables in the present study probably do not justify the effort involved in their calculation.

CHAPTER 3

ESTIMATING PREMORBID WAIS-R FACTOR SCORES

3.1. Introduction.

In view of the superior construct validity of WAIS-R factor scores (Crawford, O'Carroll & Venerri, 1998), and the robustness of the factor structure across cultures, and in the presence of neurological disease (e.g., Atkinson, Cyr, Doxey & Vigna, 1989), it is proposed in this chapter to examine the criterion validity of models to estimate factor scores from demographic information alone, from demographic information and NART in combination, and from NART alone. As in chapter 2, the potential contribution of interactions between predictor variables will be assessed statistically.

Factor IQ scores have been shown to have greater clinical utility than conventional summary IQ measures and subtest scatter indices (Crawford, Johnson, Mychalkiw & Moore, 1997). Using discriminant function analyses, these authors showed that factor scores achieved a statistically higher classification accuracy in the differentiation of closed head injury subjects from matched healthy controls. As previously noted, another incidental, but important finding with implications for the clinical interpretation of subtest profiles, was, that the healthy controls exhibited greater subtest variability than the closed head injury subjects. Although VIQ and PIQ were significantly different in the head injury group in this study, the mean difference of 2.7 points is modest compared to the discrepancy of 20.5 points obtained between the attention-concentration factor score and the verbal factor score. The subtests contributing

to the attention-concentration composite (i.e. Digit Span and Arithmetic) make little contribution to the verbal factor, consistent with the observation that the two groups were not differentiated by the verbal factor. Thus, a moderate deficit on VIQ masked an absence of deficit on the verbal factor, and a severe deficit in attention/concentration. Lezak's (1988) observation that Arithmetic, Digit Span and Digit Symbol are the subtests which most commonly reflect impairment following brain injury, provides empirical support for factorial analyses which consistently generate an attention-concentration factor consisting of principal loadings from Digit Symbol, Arithmetic and Digit Span.

In the present study the factor method of Atkinson (1991) is resorted to because of the ease of calculation of the factor-based short-forms, and because of the quality of information provided for clinical interpretation of their use. Atkinson (1991) has provided a table of standard deviations, reliability coefficients, and standard errors of estimation and prediction for the nine age bands of the WAIS-R standardisation (Wechsler, 1981). He has also provided a table of critical values of estimated-obtained factor score discrepancies for four levels of significance and a table of abnormality of factor score differences. The aim of the present study is to extend this work by Atkinson, to build equations to predict factor scores for use in the UK using demographics and NART, and removing the last barrier to their use in clinical practice.

3.2 Method and Procedure.

Factor scores were calculated for the WAIS-R data of the healthy sample (N=245) described in Chapter 2. These factors are firstly, a Verbal factor (V), comprised of Vocabulary, Information, and Comprehension; secondly, a Perceptual Organisation factor (PO), comprised of Block Design and Object Assembly; and thirdly, an Attention-Concentration/Freedom from Distractibility factor (A/C), comprised of Digit Span and Arithmetic. These combinations of subtest groupings satisfy the criteria described in the review of factor models by Leckliter, Matarazzo & Silverstein (1986) whereby groupings are justified on the basis of high and main loadings consistently across samples. A factor score is calculated by adding the scaled scores for that factor (X_o), and converting this to a deviation quotient (*DQ*) with mean 100 and standard deviation 15 according to the following formula:

$DQ = 15/SD_o(X_o - M_o) + 100,$

where SD_o = standard deviation of the observed distribution calculated with Tellegen and Briggs' (1967) formula, and M_o = mean of the obtained distribution (i.e., 30 for V, 20 for PO, and 20 for A/C).

Summary IQ scores were calculated following the standard scoring procedures in the WAIS-R manual (Wechsler, 1987).

Critical values for the evaluation of discrepancies between estimated and obtained factor scores were calculated for four levels of significance by multiplying the

corresponding z value (one-tailed) by the standard error of estimate (SE_{est}) for each of the relevant regression equations.

In order to determine if the modelling of interactions between predictor variables significantly improves prediction of variance in the criterion variables over that achieved by the Demographics and NART plus demographics models alone, the contribution of all possible two-way interactions was calculated by further regression analyses according to the procedure described in Chapter 2 for the subtest models. A series of variables were created according to the method of Cohen & Cohen (1982) by multiplying each of the predictor variables with each of the other predictor variables to represent all possible two-way interactions. These variables were then entered in separate two-stage hierarchical regression analyses following forced entry of the NART and primary demographic variables. In the case of the NART plus demographics models there were five predictor variables and therefore ten additional variables were created to represent the two-way interactions. In the case of the Demographics models the same procedure was adopted except for the exclusion of those interaction variables incorporating NART. Again, in view of the potentially large number of variables, the analyses were confined to potential two-way interactions unless these demonstrated an important effect. Interactions were also evaluated in respect of summary IQs.

3.3 Results.

Summary statistics for the sample's scores on the three factors are shown as Table 3.1.

Table 3.1: Atkinson factor scores, healthy UK sample (N=245).

Factor	Mean	Std. Dev.	Minimum	Maximum
V	100.6772	14.4406	61.68	134.67
PO	100.0605	14.6523	61.61	146.62
A/C	100.8393	14.7573	63.55	137.07

The means and standard deviations obtained on the factors correspond very closely to the ideal figures for a representative general population sample.

3.3.1: Estimating premorbid factor scores.

The summary statistics for estimated Atkinson factor scores derived from the Demographic, NART plus demographic and NART models are shown as Table 3.2.

Table 3.2: Summary statistics for estimated Atkinson factor scores: Demographic, NART plus demographics and NART models.

	Demographic models		NART + Dem. models				NART models	
Factor	R^2	SE_{est}	R^2	SE_{est}	sig*	Sig**	R^2	SE_{est}
V	.512	10.17	.680	8.25	$p<.001$	$P<.001$	.611	9.03
PO	.255	12.76	.271	12.64	$p=.021$	$P<.001$	.174	13.34
A/C	.320	12.27	.414	11.42	$p<.001$	$P<.001$	.355	11.88

Note The NART plus demographics models were created in two-stage regression analyses where demographics were entered before NART and where NART was entered before demographics; sig*= significance of change in R^2 following inclusion of NART to demographic predictor variables in regression analyses; Sig**= significance of change in R^2 following inclusion of demographic predictor variables to NART in regression analyses.

As can be seen from Table 3.2, the combination of NART and demographics is a more powerful predictor of Atkinson factor scores than the Demographic and the NART models. The NART plus demographics models were created initially by adding NART to demographic information in separate two-stage hierarchical regression analyses. This led to a generally positive increase in the variance accounted for in the criterion variables ($p<.001$ for V; $p=.021$ for PO; and $p<.001$ for A/C). Secondly, the NART plus demographics models were created with separate two-stage hierarchical regression analyses with demographics entered after NART. The variance in the criterion variables accounted for increases significantly (all p values $< .001$).

NART plus demographics estimates 68% of the variance in the verbal factor (V), 27% of the variance in the perceptual organisation factor (PO) and 41% of the variance in the attention concentration/freedom from distractibility factor (A/C). Comparison of the NART and Demographic models shows that the NART model is more successful in respect of V and, surprisingly perhaps, A/C, but the Demographic models are marginally more successful in relation to PO. Where the clinical circumstances permit it, clearly, clinicians will wish to apply the combined NART plus demographics models in estimating premorbid factor scores, as in all cases they were the most powerful set of models.

3.3.2. Demographics factor models.

The regression equations for estimating premorbid factor scores from demographic information are shown as Table 3.3.

Table 3.3: Regression equations for estimating premorbid factor scores from demographic information alone.

V = 77.108+(0.233 x age)-(3.651 x sex)+(2.393 x educ)-(3.874 x class).
PO = 97.422+(0.129 x age)-(5.132 x sex)+(1.213 x educ)-(3.601 x class).
A/C = 97.137-(0.06575 x age)-(3.422 x sex)+(1.573 x educ)-(4.069 x class).

3.3.3: NART plus demographics factor models.

The regression equations for estimating premorbid factor scores from NART and demographic information combined, are shown as Table 3.4.

Table 3.4: Regression equations for estimating premorbid factor scores from NART and demographic information combined.

V = 114.397+(0.133 x age)-(4.167 x sex)+(0.771 x educ)-(1.979 x class)-(0.879 x NART).
PO = 109.292+(0.09671 x age)-(5.296 x sex)+(0.696 x educ)-(2.998 x class)-(0.280 x NART).
A/C = 123.635-(0.01098 x age)-(3.816 x sex)+(0.333 x educ)-(.2.622 x class)-(0.672 x NART).

3.3.4: NART factor models.

The equations for estimating factor scores from NART are as presented as Table 3.6.

Table 3.6: Regression equations for estimating premorbid factor scores from NART:

V = 123.855 - (1.189 x NART).
PO = 112.614 - (0.644 x NART).
A/C = 118.902 - (0.926 x NART).

3.3.5: Estimated-obtained factor score discrepancies.

To assist in the interpretation of discrepancies between estimated and obtained factor scores, the standard error of estimate (SE_{est}) for each factor was multiplied by the appropriate value of *z* (one-tailed) for four levels of significance, and this information is shown in Table 3.7 for use when applying the Demographics models, the NART plus demographics combined models, or the NART models.

Table 3.7 Discrepancies between estimated and obtained factor scores required for four levels of significance for the Demographics Models, the NART + Demographics Models, and the NART models.

	Demographics			NART+Demographics			NART		
Sig.	V	PO	A/C	V	PO	A/C	V	PO	A/C
15%	11.10	13.14	16.05	9.02	13.02	13.02	9.30	13.74	12.23
10%	13.79	16.33	19.94	11.21	16.18	16.18	11.56	17.08	15.20
5%	17.67	20.92	25.55	14.36	20.73	20.73	14.81	22.88	19.48
1%	25.00	29.59	36.14	20.31	29.32	29.32	20.95	30.96	27.55

Note: All values are one-tailed.

To illustrate the use of this table, take the example of a patient who obtains a score of 84 on the attention-concentration/freedom from distractibility factor. The patient is male (coded 1), aged 25 years, is employed as a computer programmer (coded as social class 2), has completed 16 years of education and has a NART error score of 8. Entering these data into the equation for the A/C factor from NART plus demographics in Table 3.4 yields an estimated premorbid A/C factor score of 114.2525. The discrepancy between the estimated and obtained score (114.2525 - 84) is 30.2525. Referring to Table 3.7, it can be seen that this discrepancy is greater than the critical value which a discrepancy must exceed (29.32)

for significance at the 1% level. Less than 1% of the general population would exhibit a discrepancy of this magnitude, and the patient's score is therefore significantly impaired.

3.3.6: Factor score interactions.

Following the rationale and method in relation to subtest scores described in Chapter 2, Table 3.8 shows the effect of including interactions amongst the independent variables into the regression analysis models. This did not lead to a significant improvement in prediction of variance in any of the Atkinson factors.

Table 3.8.:Effect of interactions on estimation of variance in factors.

	Demographics			NART+Demographics		
Factor	R^{2A}	R^{2B}	Sig.*	R^{2A}	R^{2B}	Sig.*
V	.715	.733	.052	.680	.692	.506
PO	.505	.537	.085	.271	.316	.134
A/C	.565	.575	.699	.414	.422	.969

Note: R^{2A} = variance in factor estimated by primary model; R^{2B} = variance in factor estimated following inclusion of interactions; Sig.* = significance of change in R^2.

3.3.7: Estimating premorbid summary IQ scores.

For comparative purposes, summary WAIS-R IQ scores are estimated using the Demographics, NART plus demographics, and NART models. Table 3.9 shows the summary statistics from separate hierarchical regression analyses comparing the Demographics, the NART plus demographics, and the NART models. The NART plus

demographics models were created by adding NART to demographics and by adding demographics to NART.

Table 3.9: Summary statistics estimating IQ scores with Demographics, NART plus demographics, and NART models.

	Demographics		NART + Dem.				NART	
	R^2	SE_{est}	R^2	SE_{est}	Sig*	Sig**	R^2	SE_{est}
VIQ	.540	8.79	.711	6.98	$p<.001$	$p<.001$	.641	7.72
PIQ	.313	11.16	.375	10.67	$p<.001$	$p<.001$	.312	11.10
FSIQ	.519	9.28	.652	7.91	$p<.001$	$p<.001$	.571	8.70

Note: The NART plus demographics models were created in two-stage regression analyses with demographics entered before NART and with Demographics entered after NART; Sig.* = significance of change in R^2 with Demographics entered before NART, and Sig.** = significance of change in R^2 with Demographics entered after NART.

As can be seen from Table 3.9, each model is capable of estimating quite impressive variance in VIQ and FSIQ, with lesser ability in respect of PIQ. NART plus demographics are the most powerful models followed by NART alone and followed by the Demographics models. These findings are consistent with previous work estimating WAIS-R performance from demographic information and NART (e.g. Nelson & Willison, 1982; and Crawford, 1990).

The addition of NART to demographic information in the creation of the NART plus demographics models produces a significant increase in prediction of the variance

accounted for in the criterion variables. The same effect is seen where demographic information is added after NART.

3.3.8: Summary IQ interactions.

In Chapter 2 the potential effects of interactions between predictor variables to prediction were evaluated in respect of WAIS-R subtest scores. Table 3.10 shows the effect of entering interactions between the independent variables into the regression analyses. As with the Atkinson factors, this did not lead to significant increase in estimation of variance in WAIS-R summary IQ scores.

Table 3.10.: Effect of interactions on estimated variance in WAIS-R IQ scores.

	Demographics.			NART+Demographics.		
IQ scale	R^{2A}	R^{2B}	Sig.*	R^{2A}	R^{2B}	Sig.*
VIQ	.540	.556	.204	.843	.847	.847
PIQ	.313	.344	.099	.613	.641	.188
FSIQ	.519	.536	.181	.807	.815	.520

Note: R^{2A} = variance in IQ score estimated by primary model; R^{2B} = variance in IQ score estimated following inclusion of interactions; Sig.* = significance of change in R^2.

3.4. Discussion.

In this Chapter a range of models for estimating premorbid Atkinson factor scores from demographic information, from NART plus demographic information and from NART have been presented. As was found in relation to the models for estimating premorbid WAIS-R subtest scores, the NART plus demographics models were most successful in predicting variance in factor scores and summary WAIS-R IQ scores.

A range of regression equations for estimating premorbid factor scores from NART and demographic information were constructed to meet a variety of clinical circumstances.

To assist in the interpretation of discrepancies between estimated and obtained factor scores, critical values were calculated for four levels of significance for all three model types.

Interactions between pairs of predictor variables did not exert a significant effect on prediction of variance in the dependent variables, and for this reason further potential interactions were not investigated.

For comparative purposes models were constructed to estimate premorbid summary WAIS-R IQ scores, again using demographic predictor variables, NART and demographics in combination, and from NART alone. The effects of potential two-way interactions between predictor variables were evaluated and found not to significantly affect the prediction of premorbid WAIS-R summary IQ scores.

In respect of summary IQ indices, the NART plus demographics equations presented here to estimate premorbid ability predict 71%, 38% and 65% of the variance in

WAIS-R VIQ, PIQ, and FSIQ respectively. In contrast, the NART plus demographics equations presented to estimate premorbid Atkinson factor scores predict 68%, 27% and 41% of the variance in V, PO, and A/C respectively. Given the stability of the underlying ability dimensions represented by the factor models, and the relative sensitivity of *weighted* factor scores in comparison with summary IQ measures in clinical differentiation (e.g. Crawford, Johnston, Mychalkiw & Moore, 1997), the relative performance of Atkinson factor scores in differentiating between healthy and clinical cases will be of considerable interest. The intention is to establish if either type of WAIS-R summary index can be recommended for use in clinical neuropsychological practice. In addition to the superior construct validity of factor scores, testing with the Atkinson short form takes less time, allowing the clinician more time to complete other aspects of the clinical assessment. The use of individual subtest equations and factor equations to differentiate between impaired and normal subjects will be examined in Chapter 4. The utility of factor scores relative to summary IQ scores will also be evaluated, extending work previously conducted using *weighted* factor scores (Crawford, Johnston, Mychalkiw & Moore, 1997). These authors showed that factor scores outperformed summary IQ scores in differentiating between impaired head injury subjects and matched healthy controls.

Having demonstrated the statistical superiority of the NART plus demographics model compared to the Demographics and NART models, it is anticipated that the former will be most successful in clinical differentiation for most ability indices. However, there are clinical circumstances where the use of a present ability measure is contraindicated by the patient's clinical condition, and for this reason the Demographics model will be

evaluated also in the clinical sample described in Chapter 4. The relative performance of the NART model is of some interest in view of what is known about the impairment sensitivity of the NART in some clinical groups which are included in typical general neurological samples.

CHAPTER 4

DISCRIMINATION BETWEEN HEALTHY AND IMPAIRED CASES USING ESTIMATED-OBTAINED DISCREPANCY SCORES

4.1. Introduction.

The present chapter examines the ability of discrepancies between obtained scores and estimated premorbid ability scores to differentiate between cases from a large, heterogenous, neurological sample (N=298), and cases from the large healthy sample (N=245) described in Chapter 2.

In the previous two chapters, regression equations have been presented, based on a large (N=245) census-matched UK healthy sample, for use in estimating premorbid ability, using NART and demographic predictor variables. As described, the regression equations estimate premorbid WAIS-R subtest scores, summary IQ indices, and unweighted factor scores (Atkinson, 1991) from the test performance of the healthy sample presented in Chapters 2 and 3. The key issue is the utility of the regression models to differentiate between healthy and clinical cases - i.e. to assess their ability to correctly classify the individual case in clinical practice, and their ability to avoid mis-classification of healthy subjects.

The history of the quest for reliable clinical methods capable of unequivocal differentiation between healthy controls and impaired cases is littered with failed effort; notable amongst these is the study (previously cited) by O'Carroll, Curran, Ross, Murray,

Riddle, Moffoot, Ebmeier & Goodwin (1994), in which they attempted to differentiate between patients with Alzheimer's disease, patients with major depressive illness and controls using NART and WMS-R measures in a discrepancy analysis. In spite of well separated mean scores which were highly significant, overlap between groups rendered the indices unusable for clinical differentiation in the individual case.

Using their US demographic equations, Paolo, Ryan, Troster & Hilmer (1996) showed that estimated scaled scores were significantly greater than obtained scaled scores on all 11 subtests in their general clinical sample (N=247; males=221, females=26; mean age 58.12; s.d.=12), with all p values <.001.

They presented one-tailed confidence intervals at the 90% and 95% levels of significance, based on the SE_{est} for each subtest and rounded to the *nearest* whole number. For all subtests, the rounded 90% interval was defined as 3 points, and the 95% as 4 points. They state that if the obtained scaled score is greater than 3 points below the estimated score, it falls below the 10% confidence limit, i.e., the obtained discrepancy is beyond that expected on the basis of measurement error in 90% of their standardisation sample, and therefore significant.

In this study, the proportions of the US standardisation sample evidencing discrepancies on individual subtests will be compared to the incidence of discrepancies obtained in the UK healthy sample using the equations developed in Chapter 2. Given the concern of this study to evaluate the inferences which can be made on the basis of single subtest scores, the relative proportions of the healthy, and clinical samples evidencing apparent decline on one or less subtests is of considerable interest. Paolo *et al.,* (1996)

showed that at the 10% level of significance, only 16% of the standardisation sample had more than one subtest evidencing apparent decline, whereas 65% of the clinical sample had evidence of apparent decline on more than one subtest.

At the more stringent 5% level of significance, 6% of the healthy sample and 41% of the clinical sample respectively evidenced significant apparent decline on two or more subtests. However, 1% of the healthy sample evidenced apparent decline on six or more subtests.

With regard to the respective sample proportions evidencing no decline at the 5% level of significance on any subtest, 84% of the healthy sample, and 40% of the clinical sample were not discrepant respectively.

On the basis that an estimated-obtained discrepancy suggests acquired intellectual impairment, the aim of this chapter is to examine the ability of discrepancies between estimated and obtained scores to discriminate between healthy and clinical cases. The performance of the UK demographic models will be compared to the Paolo *et al.* demographic models, and the effect of incorporating NART will be evaluated.

4.2. Method.

From a database containing clinical, neuropsychological and demographic information pertaining to 510 cases referred to four UK clinical neuropsychology services (Aberdeen, London, Edinburgh, and Stoke-on-Trent) by a research group which included the present writer, a general clinical sample (N=298; males=187, females=111) was selected. The data custodian is Professor J R Crawford with joint, several and individual

ownership of the data with the members of the research group. All testable subjects from the database who had completed a full-length WAIS-R, NART, and had demographic information recorded were included in the study, provided the subject had a diagnosed neurological condition. Subjects with concussional brain injury were clinically ascertained to have emerged from PTA prior to neuropsychological assessment. Of the original sample in the database 212 subjects were excluded because of missing clinical, neuropsychological or demographic information. The breakdown of the sample by diagnostic group is shown as Table 4.1.

Table 4.1.: Breakdown of the Clinical sample (N=298) by diagnostic group.

DIAGNOSIS	N
Closed head injury	152
Dementia Alzheimer type	38
Multi-infarct dementia	8
Dementia other (e.g. post viral, anoxic, alcoholic)	38
Huntington's Disease	15
Epilepsy and intracerebral tumour	17
Subarachnoid haemorrhage	18
Stroke	7
Miscellaneous neurological	5
Total:	298

4.3. Procedure.

All subjects had been administered a full-length WAIS-R for clinical purposes following standard procedures (Wechsler, 1981), and a full-length NART (Nelson, 1982). The demographic characteristics of each subject were recorded concerning age, sex, years education and occupation. In recording years of education, subjects were credited 0.25 of a year for each whole year of attendance at day or evening classes provided this led ordinarily to a recognised qualification, in addition to their years of conventional full-time education. In coding social class, married females not in full-time employment were coded by their husband's occupation, and in all cases highest lifetime occupation was recorded to control for the effects of contemporary unemployment or the phenomenon of "winding down" to retirement. Socio-economic grouping was determined using the Classification of Occupations (OPCS, 1980). Sex was dummy variable coded (Cohen & Cohen, 1983) with males = 1, and females = 2. The mean age of the sample was 43.05 years (SD=17.34) with a range from 16 to 81. The mean years of education was 11.59 (SD=2.37) with a range from 5 to 20.

The mean age of the clinical sample reported by Paolo *et al.* (1996) is significantly greater (M=58.12 years), probably reflecting the higher proportion of cases with stroke disease in the US sample.

Estimated ability scores with respect to subtests, factor scores, and summary IQ scores were computed for the clinical sample using the Demographics, NART plus demographics, and NART regression equations derived from the healthy sample and presented in Chapters 2 and 3 of this study. Estimated minus obtained scores were

calculated for each subject for each WAIS-R variable to form new *discrepancy* variables. These discrepancy scores were then stratified using the appropriate critical discrepancy scores for four levels of significance (15%, 10%, 5%, and 1%) presented in Tables 2.8, 2.9, 3.6, and 3.9.

For comparison with the Paolo *et al.* (1996) study, the number of healthy and clinical subjects discrepant on from none to up to all 11 subtests was calculated for the UK Demographics equations, and subsequently for the NART plus demographics and the NART equations.

The test scores of the clinical cases were transformed to *z* scores using the mean and standard deviations of the relevant index in the healthy sample. Any score lower than –1.64 (one-tailed value of *z* for significance at ≥5% level) was deemed abnormal. Thus it was possible to calculate and compare the relative proportions of the clinical and healthy samples with abnormal scores on the criterion variables.

A series of hierarchical discriminant function analyses were then conducted for each subtest and for each composite index, to assess differentiation between healthy and clinical cases. In the first step, the obtained scores were entered for analysis and the relevant summary statistics generated. These included the overall percentage of cases correctly classified as either healthy (H) or impaired (C). In the second step, the estimated premorbid scores were entered to determine if these premorbid scores would significantly improve discrimination (i.e. by operating as discrepancies) between healthy and clinical samples over that achieved by the obtained scores alone.

4.4. Results.

With regard to socio-economic distribution, a chi-square test showed that the social class distribution in the clinical sample did not differ significantly from the healthy sample ($\chi^2 = 4.04$, d.f. = 4, $p = .40$), and these data are shown as Table 4.2. Mean NART error score of the healthy sample was 19.50 (SD = 9.49), and 25.74 (SD = 11.44) in the clinical sample ($t = 6.949$, d.f. = 540.95, $p < .001$), showing that the clinical sample made significantly more NART errors. The distribution of age in the two samples was surprisingly similar, and not significantly different ($t = 0.021$, d.f. = 513.324, $p = .983$). Variance in the distributions of age in the two samples was not significantly different (Independent samples test F = 0.27; $p = .869$). With regard to educational achievement, the healthy sample had significantly more years education than the clinical group ($t = 4.756$, d.f. = 4548.385, $p < .001$). Variance in the distributions of years of education in the two samples also differed significantly (Independent samples test F = 33.606; $p < .001$).

There was an excess of males in the clinical sample relative to the healthy sample ($\chi^2 = 9.20$, d.f. = 1, $p = .0024$) reflecting the generally recognised epidemiological sex imbalance in neurological impairment caused by head injury (see e.g. Jennet, Murray, Carlin, McKean, MacMillan & Strang, 1979).

Table 4.2.: Socio-economic distribution of the Healthy (N=245) and Clinical (N=298) samples.

OPCS group	5	4	3	2	1	χ^2	d.f.	p
Healthy sample	20	41	107	60	17			
						4.04	4	0.4
Clinical sample	15	56	146	64	17			

The psychometric characteristics of the clinical sample relative to the healthy sample are shown as Tables 4.3 and 4.4. The clinical sample is significantly impaired on all WAIS-R ability indices (all p values <.001) as was found in the study by Paolo *et al.*(1996).

Table 4.3. Summary statistics for the healthy (N=245) and the clinical (N=298) samples: WAIS-R subtests.

	Healthy sample			Clinical sample				
Subtest	Mean	s.d.	Range	Mean	s.d.	Range	*t*	*p*
Inf.	9.79	3.08	3-18	7.85	3.30	0-18	7.093	*
D.Span	10.83	2.77	3-19	8.19	2.88	1-16	10.818	*
Vocab.	10.05	2.52	4-18	8.55	2.88	1-18	6.387	*
Arith.	10.97	3.02	5-17	8.07	3.30	1-17	10.595	*
Comp.	9.81	2.76	2-18	8.20	3.44	1-19	6.033	*
Sim.	9.00	2.69	2-15	7.55	3.30	0-18	5.658	*
PC	8.94	2.87	3-17	7.05	3.33	1-17	7.093	*
PA	9.51	3.19	2-17	6.93	3.30	1-17	9.181	*
BD	9.63	3.03	4-19	7.61	3.40	1-17	7.231	*
OA	8.46	2.79	1-18	6.72	3.35	1-15	6.508	*
Dsym.	8.93	2.86	2-18	5.50	3.12	0-16	13.187	*

Note: *= all *p* values <.001, following Bonferonni correction where the pairwise *p* values must be <.0045 to achieve a family-wise significance at the 5% level.

Table 4.4. Summary statistics for the healthy (N=245) and the clinical (N=298) samples: WAIS-R summary indices.

	Healthy sample			Clinical sample				
	Mean	s.d.	Range	Mean	s.d.	Range	*t*	*p*
V	100.6772	14.4406	61.68-134.67	89.9991	18.2350	48.97-138.60	7.164	*
PO	100.0605	14.6523	61.61-146.62	86.9825	19.9102	44.24-149.27	8.552	*
A/C	100.8393	14.7573	63.55-137.07	82.7736	17.3228	43.40-125.23	13.120	*
VIQ	103.62	12.85	73-133	91.06	15.28	53-137	9.907	*
PIQ	102.42	13.36	67-139	86.78	15.93	57-136	12.443	*
FSIQ	103.07	13.26	71-140	88.63	15.49	50-137	11.706	*

Note: *= all *p* values <.001 following Bonferroni correction where the pairwise *p* values must be <.0083 to achieve a family wise significance at the 5% level.

4.4.1: Estimated-obtained discrepancy frequencies.

The cumulative discrepancy frequencies for each model (percent), for subtests and summary indices for four levels of significance are shown as Tables 4.5 to 4.7, and as Tables 4.9 to 4.11 respectively. These show the base rates of such discrepancies in the healthy sample as well as their frequency of occurrence in the clinical sample.

For example, with reference to Table 4.5, in the case of the Digit Symbol subtest, it can be seen that 13.5% of the healthy sample exhibited a discrepancy between their obtained subtest score and that estimated from the demographic equation which exceeded the critical value for the .15 level of significance. In contrast 69.5% of the clinical sample exhibited a discrepancy that exceeded the .15 level.

The relative proportions of the two samples exhibiting discrepancies equal to or greater than the critical values for significance at the 5% level are also shown in these Tables. These are generally highly significantly different, consistent with significant impairment in the clinical sample.

The performance of the estimator models relative to the performance of the two samples on the ability indices themselves is of considerable interest. Table 4.8 shows the percentages of subjects in the two samples exhibiting abnormal subtest scores and abnormal estimated subtest scores at the 95% confidence level and beyond. It can be seen that the performance of the two samples on the ability measures themselves is generally significantly different. Table 4.12 shows the equivalent information in respect

of the summary indices, where again performance on the ability indices themselves is significantly impaired in the clinical sample.

A primary aim of the present study is to evaluate the possibility of making inferences from a single subtest, and for this reason the base rates of cumulative frequencies of numbers of subtests discrepant in the UK healthy sample was calculated, for comparison with that found in the clinical sample. Table 4.13 shows the cumulative frequencies of number of subtests discrepant ≥5% level of significance for the Demographic models comparing the US (Paolo *et al.,* 1996) study and the present UK study.

The relative performance of the three UK models in respect of the proportions of the UK samples discrepant on up to all 11 WAIS-R subtests can potentially provide indirect evidence about the extent to which the NART may be impairment sensitive in a clinical sample. Table 4.14 shows the cumulative frequencies of number of subtests discrepant ≥5% level of significance for the Demographics, NART plus demographics, and the NART models. The data is presented in a simplified form as Table 4.15. This shows the frequency of presence or absence of subtest discrepancies at ≥5% level of significance for the three estimator models. Whilst all three models have similar classification performance within the healthy sample, within the clinical sample, the NART model has an incongruous and weaker classification performance with significantly fewer clinical cases discrepant than the other two models. The possible reasons for this are discussed below, but it appears superficially that the NART is impairment sensitive in a general clinical sample relative to demographic models. The

NART is tending to underestimate impairment relative to other models, but not in the healthy sample. The NART plus demographics model has no apparent increased sensitivity over the Demographics model in terms of its ability to measure estimated-obtained discrepancies in these healthy and clinical samples at this level of significance.

Table 4.5. Discrepancy frequencies % by subtest in the Healthy (N=245) and Clinical (N=298) samples: Demographics models.

	Healthy sample				Clinical sample					
	≥15%	≥10%	≥5%	≥1%	≥15%	≥10%	≥5%	≥1%	χ^2	*p*
Inf.	14.7	10.6	4.1	1.2	36.9	27.2	18.5	9.1	26.37	*
D.Sp.	14.7	9.4	4.9	1.6	45.9	34.2	23.5	8.4	36.25	*
Voc.	15.9	9.4	3.3	0.0	35.4	28.4	17.0	6.0	26.63	*
Arith.	16.0	8.2	3.7	0.8	57.1	49.7	40.0	13.8	98.12	*
Comp.	13.5	9.0	4.9	0.8	33.6	28.9	20.2	9.1	19.56	*
Sim.	17.6	9.0	3.3	0.4	36.9	26.2	19.2	8.1	32.10	*
PC	13.8	9.3	2.8	0.8	47.3	34.8	23.4	6.0	34.23	*
PA	12.7	8.2	4.9	0.8	41.7	35.0	25.9	10.1	36.25	*
BD	13.9	9.8	6.1	1.6	40.6	30.4	22.5	9.1	37.14	*
OA	14.3	8.6	5.3	0.4	40.6	30.9	23.2	9.8	32.49	*
DSym.	13.5	7.4	4.5	0.8	69.5	63.5	51.2	28.4	140.01	*

Note: χ^2= relative proportions of the two samples with significant discrepancies ≥ 5%; *p**<.001 following Bonferonni correction for multiple comparisons where the pairwise *p* values must be <.0045 to achieve a family wise significance at the5% level.

Table 4.6. Discrepancy frequencies % by subtest in the Healthy (N=245) and Clinical (N=298) samples: NART plus demographics models.

	Healthy sample				Clinical sample					
	≥15%	≥10%	≥5%	≥1%	≥15%	≥10%	≥5%	≥1%	χ^2	*p*
Inf.	11.4	6.5	3.2	1.2	24.5	20.8	13.8	4.7	18.03	*
D.Sp.	15.1	9.4	4.9	0.4	39.2	32.2	19.8	5.7	26.27	*
Voc.	13.8	8.1	2.4	0.4	25.5	18.4	12.4	5.7	18.32	*
Arith.	15.5	8.6	3.7	0.8	51.7	44.6	30.2	10.4	63.47	*
Comp.	11.0	9.4	5.3	0.8	29.2	23.1	12.7	6.0	9.40	.002
Sim.	13.5	9.8	5.7	0.0	27.5	22.4	14.0	5.0	10.21	.001
PC	13.0	8.7	3.2	0.8	33.2	25.0	17.1	4.0	27.52	*
PA	13.1	8.6	4.1	0.0	37.6	28.6	17.5	3.4	23.76	*
BD	11.8	10.2	5.7	2.0	38.3	31.8	22.8	7.7	30.68	*
OA	15.1	8.2	4.9	0.4	35.9	27.1	20.1	6.0	28.02	*
DSym.	12.6	7.7	4.0	1.6	68.8	60.0	49.3	24.8	133.93	*

Note: χ^2= relative proportions of the two samples with significant discrepancies ≥ 5%; *p**<.001 following Bonferonni correction for multiple comparisons where the pairwise *p* values must be <.0045 to achieve a family wise significance at the 5% level.

Table 4.7. Discrepancy frequencies % by subtest in the Healthy (N=245) and Clinical (N=298) samples: NART models.

	Healthy sample				Clinical sample					
	≥15%	≥10%	≥5%	≥1%	≥15%	≥10%	≥5%	≥1%	χ^2	*p*
Inf.	14.7	10.6	4.9	0.8	23.8	16.7	7.4	3.7	6.16	.01308
D.Sp.	15.1	6.9	5.3	0.4	40.6	31.9	18.8	4.4	22.04	*
Voc.	14.7	7.7	2.8	0.4	23.2	17.4	11.4	5.0	14.09	*
Arith.	15.1	8.2	4.9	0.0	39.3	29.9	19.1	5.0	24.54	*
Comp.	13.1	7.7	4.4	2.4	25.8	18.7	12.4	5.7	11.18	*
Sim.	15.1	9.4	3.7	0.0	26.2	18.8	12.1	5.4	13.25	*
PC	15.1	7.7	2.8	0.8	31.5	22.8	17.8	5.7	30.49	*
PA	16.7	9.7	2.4	0.4	33.9	26.7	17.0	4.2	30.78	*
BD	16.7	9.4	3.3	0.0	28.2	24.1	13.7	5.0	18.03	*
OA	15.1	7.7	2.8	0.8	33.9	25.6	18.2	5.4	31.41	*
D.Sym.	18.4	12.3	4.1	0.0	55.4	46.7	33.6	9.4	71.16	*

Note: χ^2 = relative proportions of the two samples with significant discrepancies ≥ 5% level; *p** =<.001 following Bonferonni correction for multiple comparisons where the pairwise *p* values must be <.0045 to achieve a family wise significance at the 5% level.

Table 4.8: Percentage of subjects in the Healthy (N=245) and the Clinical (N=298) samples exhibiting abnormal subtest scores at the .05 level and beyond : Subtest scores, Demographics models, NART plus demographics models, and NART models.

	Subtest scores			Demographics			NART + Demographics			NART		
	Healthy	Clinical	*p*	Healthy	Clinical	*p*	Healthy	Clinical	*P*	Healthy	Clinical	*p*
I	2.9	13.4	*	4.1	18.5	*	3.2	13.8	*	4.9	7.4	.013
DSp	4.9	28.2	*	4.9	23.5	*	4.9	19.8	*	5.3	18.8	*
V	3.7	10.7	.002	3.3	17.0	*	2.4	12.4	*	2.8	11.4	*
A	7.3	36.6	*	3.7	40.0	*	3.7	30.2	*	4.9	19.1	*
C	5.3	21.4	*	4.9	20.2	*	5.3	12.7	.002	4.4	12.4	*
S	4.1	14.8	*	3.3	19.2	*	5.7	14.0	.001	3.7	12.1	*
PC	3.3	23.5	*	2.8	23.4	*	3.2	17.5	*	2.8	17.8	*
PA	5.3	27.2	*	4.9	25.9	*	4.1	17.5	*	2.4	17.0	*
BD	1.2	18.2	*	6.1	22.5	*	5.7	22.8	*	3.3	13.7	*
OA	1.6	19.6	*	5.3	23.2	*	4.9	20.1	*	2.8	18.2	*
DSy	5.7	40.9	*	4.5	51.2	*	4.0	49.3	*	4.1	33.6	*

Note: $\chi 2$ test for independent samples, $p^*<.001$ except where stated; applying Bonferonni correction for multiple comparisons all pairwise p values must be <.0045 for family wise significance at the 5% level of significance.

Table 4.9. Discrepancy frequencies % by summary index in the Healthy (N=245) and Clinical (N=298) samples: Demographics models.

	Healthy sample				Clinical sample					
	≥15%	≥10%	≥5%	≥1%	≥15%	≥10%	≥5%	≥1%	χ^2	*p*
V	11.4	8.1	3.6	1.2	42.3	36.5	25.4	13.70	47.69	*
PO	14.7	9.8	4.5	0.8	46.7	40.0	31.80	17.80	63.09	*
A/C	10.2	4.5	1.2	0.0	52.7	42.3	30.60	9.1	80.71	*
VIQ	14.7	9.0	4.1	0.4	49.3	42.6	31.50	15.1	65.49	*
PIQ	14.7	9.0	5.7	0.4	58.7	51.0	42.3	20.5	93.96	*
FSIQ	15.1	10.2	4.9	0.8	61.1	50.7	41.3	20.8	95.25	*

Note: χ^2= relative proportions of two samples with significant discrepancies at ≥5% level; *p**=<.001 following Bonferonni correction for multiple comparisons where the pairwise *p* values must be <.0083 for family wise significance at the 5% level.

Table 4.10. Discrepancy frequencies % by summary index in the Healthy (N=245) and Clinical (N=298) samples: NART plus demographics models.

	Healthy sample				Clinical sample					
	≥15%	≥10%	≥5%	≥1%	≥15%	≥10%	≥5%	≥1%	χ^2	p
V	11.4	7.7	3.2	1.2	35.6	28.2	19.8	10.7	33.98	*
PO	14.3	10.6	4.5	0.4	44.6	36.3	27.9	14.8	51.27	*
A/C	13.1	7.8	3.3	0.4	52.0	44.7	31.9	12.4	71.62	*
VIQ	13.9	9.0	4.1	0.8	47.3	37.3	29.6	13.8	58.87	*
PIQ	15.5	11.4	4.5	0.8	56.7	47.3	36.2	16.1	122.61	*
FSIQ	15.5	11.0	5.3	0.8	57.0	46.7	37.6	18.8	79.05	*

Note: χ^2= relative proportions of the two samples with significant discrepancies at ≥5% level; p*=<.001 following Bonferonni correction for multiple comparisons where the pairwise p values must be <.0083 for family wise significance at the 5% level.

Table 4.11. Discrepancy frequencies % by summary index in the Healthy (N=245) and Clinical (N=298) samples: NART Models.

	Healthy sample				Clinical sample					
	≥15%	≥10%	≥5%	≥1%	≥15%	≥10%	≥5%	≥1%	χ^2	*p*
V	15.5	10.2	4.1	0.8	31.2	24.8	18.8	9.1	27.25	*
PO	15.9	9.3	2.0	0.0	38.3	30.5	20.8	11.4	43.77	*
A/C	15.5	10.2	4.5	0.4	49.3	44.3	31.5	12.4	63.09	*
VIQ	14.3	8.9	3.2	0.8	38.9	31.5	24.5	10.7	47.75	*
PIQ	18.0	10.6	4.1	0.0	51.3	41.9	28.5	11.1	55.65	*
FSIQ	16.7	9.4	5.3	0.8	47.3	39.3	28.9	12.4	50.03	*

Note: χ^2= relative proportions of the two samples with significant discrepancies at ≥5% level; *p**=<.001 following Bonferonni correction for multiple comparisons where the pairwise *p* values must be <.0083 for family wise significance at the 5% level.

Table 4.12: Percentage of subjects in the Healthy (N=245) and the Clinical (N=298) samples exhibiting abnormal summary scores at the .05 level and beyond: Index scores and estimator models.

	Index scores			Demographics			NART+ Demographics			NART		
	Healthy	Clinical	*p*	Healthy	Clinical	*P*	Healthy	Clinical	*p*	Healthy	Clinical	*p*
V	6.5	25.5	*	3.6	25.4	*	3.2	19.8	*	4.1	18.8	*
PO	3.3	29.5	*	4.5	31.8	*	4.5	27.9	*	2.0	20.8	*
A/C	4.5	37.2	*	1.2	30.6	*	3.3	31.9	*	4.5	31.5	*
VIQ	4.1	26.5	*	4.1	31.5	*	4.1	29.6	*	3.2	24.5	*
PIQ	5.3	39.9	*	5.7	42.3	*	4.5	36.2	*	4.1	28.5	*
FSIQ	4.5	36.6	*	4.9	41.3	*	5.3	37.6	*	5.3	28.9	*

Note: *p**<.001; using Bonferonni correction, the pairwise *p* values must be <.0083 for family wise significance at the 5% level.

Table 4.13: Cumulative frequency of number of subtests discrepant at ≥5% level of significance US and UK Demographics Models.

No. of Subtests Discr.	US study (Paolo *et al.*, 1996)				Present (UK) study			
	Healthy sample (N=2010)		Clinical sample (N=247)		Healthy sample (N=245)		Clinical sample (N=298)	
	N	%	*N*	%	*N*	%	*N*	%
0	1690	84	99	40	176	71.8	69	22.9
1	201	94	46	59	44	89.8	63	44.4
2	56	97	25	69	12	94.7	45	59.4
3	28	98	18	76	7	97.6	30	69.4
4	13	99	18	83	2	98.4	28	78.8
5	4	99	15	90	4	100.0	18	84.8
6	6	99	10	94			9	87.8
7	2	99	4	95			5	89.5
8	4	99	5	97			8	92.2
9	6	100.0	4	99			5	93.4
10			1	99			10	97.3
11			2	100.0			8	100.0

Table 4.14: Cumulative frequency of number of subtests discrepant at ≥5% level of significance: UK Demographics, NART plus demographics, and NART models.

No.of Subtests Discrep.	Demographics				NART plus demographics				NART			
	Healthy sample		Clinical sample		Healthy sample		Clinical sample		Healthy sample		Clinical sample	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
0	176	71.8	69	22.9	170	69.4	71	23.8	175	71.4	113	37.9
1	44	89.8	63	44.4	49	89.4	79	50.4	50	91.8	74	62.7
2	12	94.7	45	59.4	17	96.3	54	68.2	14	97.6	36	74.7
3	7	97.6	30	69.4	4	98.0	22	75.7	2	98.4	23	82.6
4	2	98.4	28	78.8	4	99.6	26	84.3	3	99.6	13	86.9
5	4	100.0	18	84.8	1	100.0	13	88.7	1	100.0	12	90.8
6			9	87.8			8	91.4			6	92.9
7			5	89.5			9	94.5			3	93.9
8			8	92.2			5	96.2			8	96.6
9			5	93.4			4	97.6			5	98.3
10			10	97.3			3	98.6			4	99.7
11			8	100.0			4	100.0			1	100.0

Table 4.15 Frequency of presence or absence of subtest discrepancies ≥5% level by model.

	Healthy sample (N=245) Discrepancies %		Clinical sample (N=298) Discrepancies %	
	0	≥ 1	0	≥ 1
Demographics	71.8	28.2	22.9	77.1
NART+demo.	69.4	30.6	23.8	76.2
NART	71.4	28.6	37.9	62.1

4.4.2: Discriminating between healthy and clinical cases using estimated premorbid scores.

In this section the results of a series of hierarchical discriminant function analyses are presented showing the classification rates achieved in the two UK samples using WAIS-R scores, and estimated premorbid scores to establish if these operate systematically as markers of clinical impairment in the form of discrepancies. Differences between estimated and obtained scores would be expected to operate as discrepancy variables.

The results of the hierarchical discriminant function analyses for subtests are presented as Tables 4.16, 4.17, 4.18, and 4.19. In each case, the first discriminant function is the subtest score showing predicted group membership for healthy subjects and clinical cases. Subsequent discriminant functions were computed in respect of the relevant estimated subtest score. Where the relevant F to enter criterion was met, the

relevant changes in discrimination are shown. The results of the hierarchical discriminant function analyses for the unweighted factor scores are shown as Table 4.20, and those for summary IQs are shown as Table 4.21.

To illustrate, with reference to Table 4.16, it can be seen that none of the Digit Span estimated premorbid scores made any difference to the discriminant analysis (i.e. the appropriate F to enter criterion was not met. Of the healthy sample, 70.2% are correctly classified as healthy (H), and 29.8% are mis-classified as clinical or impaired (C). Of the clinical sample, 69.5% are correctly classified as clinical or impaired (C) and 30.5% mis-classified as healthy (H). The overall percentage of correctly classified cases is provided as 69.8%. In general, with regard to the WAIS-R subtests the gains in classification do not appear to justify the effort involved in the computation of the estimated premorbid variables.

With regard to the WAIS-R factor scores, reference to Table 4.20 shows that the attention-concentration/freedom from distractibility factor is the most successful index in classification. Some estimated premorbid variables make significant but modest contributions to discrimination. The WAIS-R summary IQ measures are marginally superior to the factor scores in overall correct classification.

Table 4.16: Discrimination between healthy (N=245) and clinical (N=298) samples using subtest score, and estimated subtest score (Demographics, NART plus demographics and NART models).

Function	Wilks' Lambda	Canonical Correlation	Predicted group Healthy sample %		Predicted group Clinical sample %		Overall % correct.
			H.	C.	H.	C.	
Information	.916	.290	66.9	33.1	41.9	58.1	62.1
I+ID	-	-	-	-	-	-	-
I+IDN	-	-	-	-	-	-	-
I+IN	.906	.306	63.7	36.3	38.3	61.7	62.6
D.Sp.	.822	.422	70.2	29.8	30.5	69.5	69.8
D.Sp+D.SpD	-	-	-	-	-	-	-
D.Sp+D.SpDN	-	-	-	-	-	-	-
D.Sp.+D.SpN	-	-	-	-	-	-	-
Voc.	.930	.264	60.8	39.2	33.2	66.8	64.1
Voc.+VocD	-	-	-	-	-	-	-
Voc.+VocDN	.920	.283	61.6	38.4	35.9	64.1	63.0
Voc.+VocN	.916	.290	65.3	34.7	39.6	60.4	62.6

Table 4.17: Discrimination between healthy (N=245) and clinical (N=298) samples using subtest score and estimated subtest score (Demographics, NART plus demographics, and NART models).

Function	Wilks' Lambda	Canonical Correlation	Predicted group Healthy sample		Predicted group Clinical sample		Overall % correct
			H.	C.	H.	C.	
Arith.	.828	.415	67.8	32.2	32.6	67.4	67.6
Arith+ArithD	.799	.448	69.8	30.2	29.9	70.1	70.0
Arith+ArithDN	.813	.433	66.9	33.1	29.9	70.1	68.7
Arith+ArithN	-	-	-	-	-	-	-
Comp.	.939	.247	55.9	44.1	33.3	66.7	61.8
Comp+CompD	-	-	-	-	-	-	-
Comp+CompDN	-	-	-	-	-	-	-
Comp+CompN	.914	.293	66.9	33.1	40.8	59.2	62.7
Sim.	.946	.232	59.2	40.8	37.2	62.8	61.1
Sim+SimD	-	-	-	-	-	-	-
Sim+SimDN	.938	.248	62.0	38.0	41.3	58.7	60.2
Sim+SimN	.915	.291	64.5	35.5	40.9	59.1	61.5

Table 4.18: Discrimination between healthy (N=245) and clinical (N=298) samples using subtest score and estimated subtest score (Demographics, NART plus demographics, and NART models).

Function	Wilks' Lambda	Canonical Correlation	Predicted group Healthy sample %		Predicted group Clinical sample %		Overall % correct
			H.	C.	H.	C.	
PC	.917	.288	67.3	32.7	43.3	56.7	61.5
PC+PCD	-	-	-	-	-	-	-
PC+PCDN	-	-	-	-	-	-	-
PC+PCN	.890	.332	62.4	37.6	37.6	62.4	62.4
PA	.865	.367	57.1	42.9	27.2	72.8	65.7
PA+PAD	.857	.379	64.1	35.9	29.5	70.5	67.6
PA+PADN	-	-	-	-	-	-	-
PA+PAN	.851	.386	64.9	35.1	31.2	68.8	67.0
BD	.912	.297	61.2	38.8	37.7	62.3	61.8
BD+BDD	.886	.338	66.5	33.5	35.0	65.0	65.7
BD+BDDN	.899	.318	61.2	38.8	34.0	66.0	63.8
BD+BDN	.883	.342	64.5	35.5	37.7	62.3	63.3

Table 4.19: Discrimination between healthy (N=245) and clinical (N=298) samples using subtest scores and estimated subtest scores (Demographics, NART plus demographics, and NART models).

Function	Wilks' Lambda	Canonical Correlation	Predicted group Healthy sample %		Predicted group Clinical sample %		Overall % correct
			H.	C.	H.	C.	
OA	.929	.266	60.8	39.2	41.2	58.8	59.7
OA+OAD	.921	.282	65.3	34.7	41.6	58.4	61.6
OA+OADN	-	-	-	-	-	-	-
OA+OAN	.893	.327	63.7	36.3	39.9	60.1	61.7
DSym.	.756	.494	69.0	31.0	23.6	76.4	73.0
DSym+DsymD	.689	.557	77.6	22.4	24.0	76.0	76.7
DSym+DsymDN	.701	.547	76.3	23.7	24.3	75.7	76.0
DSym+DsymN	.749	.501	68.6	31.4	25.7	74.3	71.7

Table 4.20: Discrimination between healthy (N=245) and clinical (N=298) samples using Factor scores and estimated Factor scores (Demographics, NART plus demographics, and NART models).

Function	Wilks' Lambda	Canonical Correlation	Predicted group Healthy sample %		Predicted group Clinical sample %		Overall % correct
			H.	C.	H.	C.	
V	.907	.305	65.3	34.7	36.2	63.8	64.5
V+VD	.900	.317	66.1	33.9	38.6	61.4	63.5
V+VDN	-	-	-	-	-	-	-
V+VN	-	-	-	-	-	-	-
PO	.881	.345	66.5	33.5	37.2	62.8	64.5
PO+POD	.874	.354	66.1	33.9	34.9	65.1	65.6
PO+PODN	-	-	-	-	-	-	-
PO+PON	.864	.369	67.3	32.7	38.3	61.7	64.3
A/C	.764	.486	69.0	31.0	27.5	72.5	70.9
A/C+A/CD	.737	.513	77.1	22.9	28.5	71.5	74.0
A/C+A/CDN	.743	.507	74.7	25.3	28.2	71.8	73.1
A/C+A/CN	-	-	-	-	-	-	-

Table 4.21: Discrimination between healthy (N=245) and clinical (N=298) samples using WAIS-R IQ scores and estimated WAIS-R IQ scores (Demographics, NART plus demographics, and NART models).

Function	Wilks' Lambda	Canonical Correlation	Predicted group Healthy sample %		Predicted group Clinical sample %		Overall % correct
			H.	C.	H.	C.	
VIQ	.851	.386	63.7	36.3	31.9	68.1	66.1
VIQ+VIQD	.825	.418	68.6	31.4	30.5	69.5	69.1
VIQ+VIQDN	.835	.407	66.5	33.5	29.9	70.1	68.5
VIQ+VIQN	-	-	-	-	-	-	-
PIQ	.783	.465	71.8	28.2	28.5	71.5	71.6
PIQ+PIQD	.771	.479	72.7	27.3	30.2	69.8	71.1
PIQ+PIQDN	.778	.472	70.6	29.4	28.9	71.1	70.9
PIQ+PIQN	-	-	-	-	-	-	-
FIQ	.803	.444	69.8	30.2	30.2	69.8	69.8
FIQ+FIQD	.772	.478	78.8	21.2	28.2	71.8	75.0
FIQ+FIQDN	.777	.473	75.1	24.9	29.2	70.8	72.7
FIQ+FIQN	-	-	-	-	-	-	-

4.5 Discussion.

The UK Demographics equations, using slightly different independent variables from those used in the US study, seem to overestimate discrepancy in the healthy sample, relative to the US equations (28% versus 16% at the ≥5% level of significance). The US standardisation sample employed by Paolo *et al.* (1996) to generate their equations is extremely large (N=2010), but the same trend is seen in the clinical samples where the US equations estimated discrepancies in 60% of the sample (N=247), whilst the UK equations estimated discrepancies in 77.1% of the sample (N=298). Examination of the two clinical samples shows that the UK sample had generally higher mean subtest scores, and it is therefore unlikely that the sample was more impaired than that recruited for the US study.

With regard to the demographic characteristics of the two samples, the US sample was significantly older (mean age 58.12 versus 43.05) reflecting the relative preponderance of patients with stroke disease in the sample. Similarly, the UK sample had a relative preponderance of closed head injury subjects. However, it does not seem likely that sampling characteristics (e.g. severity of impairment) could account for differences between the two demographic models particularly in view of the careful stratification of the two *healthy* samples where the effect is also present. Although the UK sample is much smaller than the US sample, it is still of considerable size and representative of the UK healthy population, in spite of it having been drawn from one

geographical area (Aberdeen). The sample contained few subjects from ethnic minorities as these are not substantially represented locally. Similarly, the US sample had a relative dearth of females and non-whites. The UK samples were more uniformly similar, except for years of education.

The addition of NART to demographic information has little effect in changing the sensitivity of detection of discrepancy in the UK clinical sample at ≥5% level of significance. The NART models however produce highly discrepant results within the clinical sample compared to the other models. In the clinical sample, 15% and 14% more individuals respectively are classified as non-discrepant by the NART models compared to the Demographics, and the NART plus demographics models. The most likely explanation of this effect is that the NART is impairment sensitive to a degree. That this should be so in a heterogenous neurological sample is not surprising in view of the studies reviewed in Chapter 1, and summarised by O'Carroll (1995), showing that the NART cannot be relied upon as a valid estimate of premorbid ability in a variety of neurological conditions, and in common neurological conditions as patients become more severely impaired. The degree to which NART appears to underestimate premorbid ability in the general clinical sample - representative as it is of that which comes through the door for assessment in typical clinical services - is, however, disappointing. All the more reason, therefore, to avoid the procedure advocated by McKinlay and Gray (1992) where the subtest comparator is a NART derived summary premorbid IQ index.

In attempting to classify an individual patient's performance on an individual subtest relative to their estimated subtest score, clinicians must bear in mind that in the

present study, approximately 10% of the healthy sample exhibited at least one discrepancy at ≥5% level of significance. Conversely, between 34% and 65% of the clinical sample exhibited no discrepancies whatever when given a *full-length* WAIS-R, depending on the particular estimator model employed.

Critical discrepancy scores for clinical evaluation of estimated-obtained discrepancies at four levels of significance have been presented for subtests and for factor scores. These provide the clinician with a range of robust models for use in a variety of clinical contexts. These permit the clinician to calculate estimated scores using equations which employ demographic information and NART scores alone and in combination.

Considering the number of subtests discrepant, shown in Table 4.14, around 95% of the healthy sample were discrepant on two or less subtests on all three models, in distinction to the frequency of occurrence of discrepancies on two or less subtests of from 25 to 40% in the clinical sample, depending on the particular estimator model employed.

To qualify as a valid means of estimating premorbid ability however, an estimator model must significantly improve our ability to detect impairment over the use of impairment-sensitive measures alone (Crawford, 1992). The data showing the percentages of the two samples exhibiting abnormal subtest scores and abnormal estimated subtest scores at the 95% confidence level and beyond via the three estimator models are shown as Table 4.8; It can readily be seen that the subtest measures alone are effective in detecting impairment, obviating the necessity to resort to the estimator

models in general. This test of the *utility* of the regression approach has not been resorted to previously.

Discriminant function analyses were resorted to in order to establish the relative ability of individual ability indices to discriminate between healthy and clinical cases, and to determine if estimated scores significantly improve our ability to discriminate when combined in discriminant functions. These data are shown in Tables 4.16 to 4.19 for the subtests; Table 4.20 for the Factor scores; and Table 4.21 for the summary IQ scores.

Surprisingly, it was shown that estimated scores had in most cases only a marginal effect on discrimination between healthy and clinical cases over that achieved by the ability indices alone. In spite of their impressive performance in terms of estimating variance in premorbid ability, the regression equations appear to be no more useful than the ability measures themselves. These findings are counter-intuitive in respect of the rationale underlying this thesis whereby it would be expected that controlling e.g. for premorbid ability should improve the discrimination between healthy and clinical cases over that achieved by the use of ability measures alone. The following example shows the effect of attempting to control for a potential source of error variance in one of the criterion variables examined in this thesis in attempting to differentiate between clinical and healthy cases.

Fig 1. Illustration of premorbid estimates acting as suppressor variables: example features the demographic estimate of premorbid performance on the WAIS-R Attention/Concentration factor.

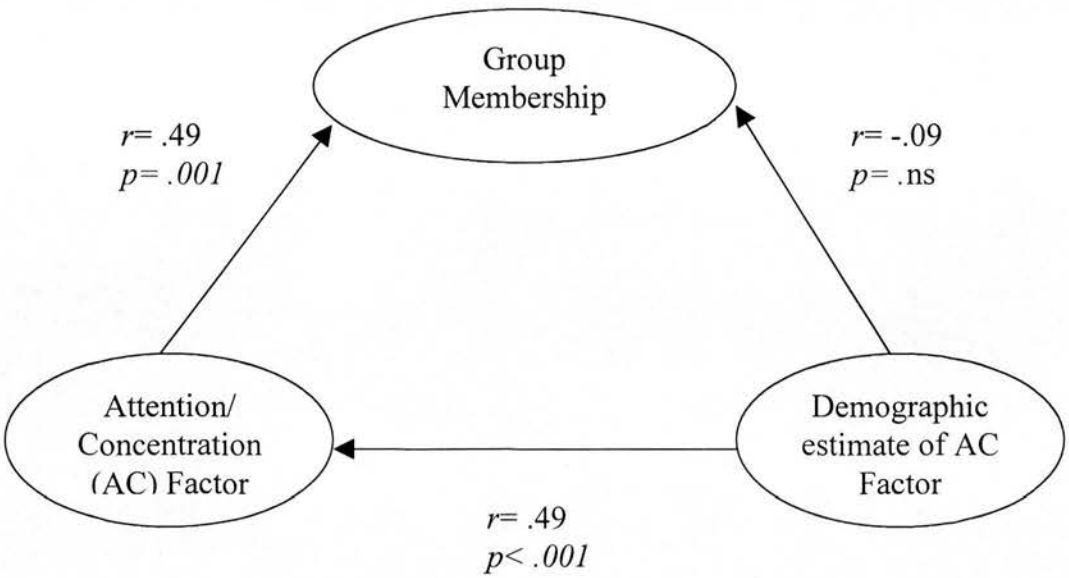

Figure 1 illustrates an example of a premorbid estimate acting as a suppressor variable. In the example shown, it can be seen that the demographic estimate of the Attention/Concentration factor is highly significantly correlated with the WAIS-R

Attention/Concentration factor (r=.49; p<.001). The latter is significantly correlated with group membership (r=.49; p<.001), whilst the former is not (r=-.09; p=ns). Partialling out the contribution of the demographic estimate of Attention/Concentration should lead to an increase in the partial correlation between group membership and the WAIS-R A/C factor. Surprisingly, only a modest increase in the partial correlation between group membership and the WAIS-R Attention/Concentration factor (to r=.5068) was found.

Clinicians are less concerned with the statistical separation of group means than they are with the need to avoid Type 1 & 2 errors in differential diagnosis in the individual case. This point is made forcefully by O'Carroll, Curran, Ross, Murray, Riddle, Moffoot, Ebmeir & Goodwin (1994) in their study comparing the performance of patients with probable Alzheimer's disease, patients with depressive illness and control subjects. In spite of well-separated mean scores, the overlap between pairs of groups prevented the use of their discrepancy quotients for valid, unequivocal clinical differentiation.

This study has shown that discrepancy scores derived from models to estimate premorbid performance do not permit unequivocal discrimination between healthy and impaired cases. In spite of highly significant differences between the WAIS-R ability indices used, all regression models mis-classified significant numbers of impaired and healthy cases.

In addition, NART has been shown to be impairment sensitive in a *heterogenous* clinical sample, and that it is appropriate therefore to use it with great care when estimating premorbid ability.

CHAPTER 5

GENERAL DISCUSSION

One of the major aims of this study has been to evaluate the criterion validity of models incorporating demographic information and NART to explain variance in WAIS-R subtest scores, Atkinson (1991) factor scores and summary IQ scores. As Crawford (1992) has described, a current ability measure must fulfil four criteria if it is to qualify as a valid means of estimating premorbid ability: Firstly, it must be reliable; secondly, it must correlate highly with IQ (criterion validity); thirdly, it must be resistant to the effects of CNS injury and disease (robustness); and fourthly, it must improve our ability to detect impairment over the use of impairment measures alone. This research has evaluated the utility of a number of predictor models incorporating NART (as a putative valid present ability measure) and demographic information. Quite surprisingly, although the predictor models were capable of explaining substantial variance in WAIS-R performance, the models did not substantially improve on the discrimination between the healthy and impaired cases achieved by the WAIS-R test data alone.

As Crawford (1992), O'Carroll (1995) and Franzen *et al.* (1997) have described, the NART has high reliability, high construct and criterion validity with variable robustness, with more information being accumulated regarding it's limitations in this regard in common clinical practice (e.g., O'Carroll, 1995). It has been shown to be impairment sensitive in the large clinical sample employed in the present study, which is broadly representative of typical clinical cases seen in the four clinical neuropsychology departments which contributed cases to the database.

Crawford (1992) has suggested that in practice by the time NART performance has deteriorated significantly, the diagnosis should not be in doubt, obviating the need for diagnostic assessment. However, in the experience of the present writer, it may not always be obvious that a patient has a clinically significant underlying dysphasic language disturbance affecting NART performance, and it would, in these circumstances, lead to an inadvertent underestimation of premorbid ability. It is suggested that the regression equations developed by Crawford, Allan, Cochrane & Parker (1990) to predict NART from demographic information be resorted to as a matter of course in clinical practice.

It is widely assumed that a hallmark of acquired intellectual impairment is increased dispersion of WAIS-R subtest scores. However, as noted, Crawford, Johnston, Michalkiw & Moore (1997) showed that subtest scatter indices were weaker than WAIS-R summary indices in differentiating between cases and healthy subjects with the scatter indices performing little better than chance. Not only were the scatter indices poor in differentiation, it was found that scatter was higher in the healthy sample.

In Chapter 4 data is presented showing the base rates of predicted-obtained discrepancies at four levels of significance in the healthy sample for estimator models using NART and demographic information alone and in combination. These models are designed to cover the range of clinical circumstances one is likely to encounter. The present research has shown that subtest discrepancies are significantly more common in clinical cases than in healthy subjects as discussed in Chapter 4, but it was also found that between 34% and 65% of the clinical sample exhibited no discrepancies whatever

when given a full-length WAIS-R depending upon the model employed to estimate premorbid subtest scores. Conversely, four subjects from the healthy sample were discrepant on five subtests using the Demographics estimator models.

Another aim of the present study was to evaluate our ability to make inferences from single WAIS-R subtests. This is a reasonable concern as there may be many circumstances in which either a full-length or recognised short-form cannot be administered. It is reassuring to note that individual WAIS-R subtests had equivalent overall correct classification rates when compared with the WAIS-R summary composite indices (Tables 4.16 to 4.21).

To assist in the interpretation of individual estimated-obtained subtest discrepancies, critical values are provided in Chapter 2 for four levels of significance, estimating from demographic information, NART plus demographics and NART alone.

If the objective of an assessment is to establish whether or not a patient is impaired, a full-length WAIS-R is probably unnecessary, but if the object of the assessment is to look for the profile of strengths and weaknesses for planning rehabilitation, then the full-length WAIS-R may be more appropriate.

The regression equations developed in the present study were capable of explaining a substantial proportion of the variance in subtest scores, factor scores and IQ scores. Incorporating two-way statistical interactions between predictor variables into the regression models did not lead to significant improvement in prediction sufficient to justify their computation.

The UK demographic equations were marginally more effective than the US equations presented by Paolo *et al.* (1996) in explaining variance in subtest scores. The addition of NART to demographic information produced substantial improvements in the proportions of variance in subtest scores it was possible to explain, consistent with previous research combining NART and demographic information in the UK (e.g., Crawford *et al.,* 1989).

It was, therefore, surprising that discrepancies between estimated premorbid and obtained WAIS-R scores did not significantly improve discrimination between healthy and clinical samples over that which could be achieved with reference to the WAIS-R obtained scores alone.

In order to investigate the potential utility of estimated-obtained subtest discrepancies relative to conventional normative and ipsative procedures for evaluating subtest scores, Paolo *et al.* (1996) stratified the test scores of their clinical sample using normative and ipsative procedures to determine whether a subtest score was relatively a weakness, a strength, or an average with respect to ±3 scaled score points in relation to the population (normative) or individual (ipsative) mean score. They then examined the relationships between these stratifications and the presence or absence of discrepancies between estimated premorbid and obtained subtest scores. They did find a relationship between "weakness" scores and discrepancies, most particularly in the ipsative procedure. The relationships between "average" and "strength" scores and the incidence of discrepancies was much weaker, especially using the normative procedure. They suggest that it may be important to take level of ability into account when evaluating

differences between estimated and obtained scores. In addition, they suggest (on the basis of the occurrence of significant discrepancies in cases with normative and ipsative "average" and "strength" scores) that the analysis of estimated-obtained discrepancies can provide additional information about subtest performance and possible intellectual impairment not always evident from normative and ipsative comparison methods.

The fundamental problem with this approach however is the large number of clinical cases with no subtest discrepancies and the incidence of healthy cases with significant discrepancies as was found in the present study.

The present study did not replicate this component of the Paolo *et al.* (1996) study, but concentrated instead on an evaluation of the effects of adding estimated premorbid scores to obtained scores in discriminating between healthy and clinical cases.

As noted, previous work comparing the performance of weighted factor scores and summary WAIS-R IQ scores in discriminating between impaired and healthy cases (Crawford, Johnston, Michalkiw & Moore, 1997) and the superior construct validity of factor scores (Crawford, O'Carroll & Venerri, 1998) led the present writer to assess the comparative ability of Atkinson-type factor scores and WAIS-R summary IQs within the same samples. Although it was found that the Attention-Concentration/Freedom from Distractibility factor was marginally the most powerful factor, it was similar in performance to the summary IQ composites. One could argue that the former are to be preferred since they involve less effort by patient and clinician. Factor scores are logically better placed than the WAIS-R composites to discriminate between cases with

lateralised brain lesions, but the main point to note is that in the present study, all WAIS-R ability indices mis-classified clinical and healthy cases.

At the time of writing this thesis, the replacement for the WAIS-R, the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III, Wechsler, 1997a) has been published in conjunction with a replacement for the WMS-R, the Wechsler Memory Scale-Third Edition (WMS-III, Wechsler, 1997b). These test batteries are designed to be used in conjunction. Provided with the tests is an excellent technical manual by Tulsky, Zhu & Ledbetter (1997), which meets many of the criticisms levelled at the WAIS-R Manual (Wechsler, 1981), discussed further below.

The justification for the replacement is based on the observation of modest gains in Wechsler IQ between standardisations of the WAIS-R and the WAIS-III: these are 1.2, 5.3 and 2.9 IQ points for VIQ, PIQ and FSIQ respectively (Tulsky, Zhu & Ledbetter, 1997; Table 4.1 p. 79). The WAIS-R now overestimates IQ. Consistent with previous work (e.g., Matarazzo, 1972), the gains are greatest in fluid ability (Performance scales).

Other advantages of the WAIS-III relative to it's antecedents include extension of the age range from 16 to 89, modernisation of the test items, extension of the floor and thereby extending the utility of the scale in quantifying deterioration, and (controversially perhaps) inclusion of mental defectives in the US standardisation sample. The tests are less reliant on timed performance, and there is more emphasis on the measurement of fluid reasoning. Although the adherence to summary Verbal, Performance and Full Scale IQs continues, the theoretical basis of the tests are strengthened by the introduction of factorially derived index scores. There is also

impressive data on reliability and validity, together with data on the relationship between test performance on the WAIS-III and WMS-III with other measures of cognitive functioning including the Standard Progressive Matrices (Raven, 1982).

As previously noted, where clinicians make interpretations on the basis of variations in score profiles they must be aware that such in-built discrepancies develop naturally over time. Thus, there is a requirement for periodic restandardisation.

The technical manual includes information on normative comparison standards, and it is intended to publish the Wechsler Verbal Fluency Test (a single word reading test similar to NART) to enable the computation of individual comparison standards. The statistical information provided in the technical manual enables the important distinction between reliable and abnormal differences to be made, on the basis of good quality information on the base rates of reliable subtest, summary scale or index deviations obtained in the healthy standardisation sample.

The potential pitfalls in over inference have been discussed in relation to Table 13 of the WAIS-R manual, which could be used for pairwise comparisons between subtests based on reliable differences without controlling for Type 1 errors, inflated by the large number of paired comparisons inherent in the method. The unwary clinician may be lured into *post-hoc* inferences thrown up by the statistical method. Tables are provided with the WAIS-III based on the ipsative method proposed by Silverstein (1982; 1984) allowing the clinician to estimate the reliability and abnormality of subtest deviations from the subject's mean score. Crawford (1999, personal communication) is developing a computer programme to extend the approach with short forms.

As with the WAIS-R, interpretation of WAIS-III performance can be conducted at the level of subtests, Summary IQs or Indices (factorially derived composite statistics). Four indices are proposed.

The index scores are factorially derived subtest groupings as follows:

- Verbal Comprehension (VCI); loadings are on Vocabulary, Similarities, and Information.
- Perceptual Organisation (POI); loadings are on Picture Completion, Block Design and Matrix Reasoning.
- Working Memory (WMI); loadings are on Arithmetic, Digit Span and Letter-Number Sequencing.
- Processing Speed (PSI); loadings are on Digit Symbol-Coding and Symbol Search.

Data on pairwise comparisons for discrepancies between indices are provided, but are not Bonferroni corrected. The indices are superior to IQs in terms of construct validity, and the reliability of the composites approaches that of the summary IQs. Factor scoring is now built into the WAIS-III so there is no lengthy calculation needed to obtain the index scores. The technical manual contains useful data on reliable differences between index scores.

Methods for estimating premorbid scores for the WAIS-III are not yet available, but this work is underway. Preliminary work from the UK standardisation has shown that ability dimensions and basic psychometric properties of the WAIS-III and WMS-III appear to be similar in the UK and US (Wechsler, 1997a).

Existing regression equations for estimating premorbid ability for WAIS-III IQs and Index scores are not applicable because of fundamental changes in the content of the new test battery. In addition, equations should be applicable to the contemporary population in view of the phenomenon of IQ gain.

Preliminary examination of criterion validity coefficients for the NART and the Wechsler Verbal Fluency Test (Form A) based on a sample of 50 healthy subjects shows that they are very similar, and that the magnitude of the coefficients for NART was similar to results for WAIS-R (Crawford, 1999, personal communication).

Once premorbid estimates are available, it will be of considerable interest to determine their ability to discriminate between healthy and impaired subjects.

The principal disadvantage of the new tests is that they are even longer than the tests which they replace, and some of the material which was not suitable for elderly subjects with poor visual acuity has not been upgraded. It remains to be seen if clinicians will abandon the other tests with which they are familiar in favour of using the WAIS-III and WMS-III in combination. There will have to be a clear clinical advantage arising from so doing. There will be many circumstances in which it will not be possible or desirable for patients to complete these quite extensive protocols. There will still be a requirement to make inferences from incomplete information, and the computer programme being developed by Crawford (1999, personal communication) will be of considerable value to those wishing to use short forms.

It would be of considerable interest to combine the Spot-The-Word Test as a potentially powerful but insensitive predictor with an impairment-sensitive criterion

measure such as Raven's Progressive Matrices (Raven, 1982). The latter has the advantage of being much shorter than the Wechsler IQ scales, and it is known to have a very high loading on *g* (e.g. Vernon, 1983). As noted, the robustness of the Spot-The-Word Test relative to the NART in a longitudinal design in a patient sample with progressive neurological disease would be of considerable interest. The study, previously cited, by Schmand, Geerlings, Jonker & Lindeboom (1998) showed that DART (Dutch NART) declined with deterioration in MMSE, and by implication, the DART would provide an underestimate of premorbid IQ in other than mild dementia.

Almost invariably, however, clinicians will be interested in mnestic performance in clinical assessment, and the disadvantage of using tests other than the Wechsler scales is that they are normalised on different populations. The clinician then has the problem of mapping between different standardisation samples. The authors of the WAIS-III and WMS-III claim that the relationship between overall intellectual ability and memory is sufficiently understood to be able to define discrepancies in mnestic performance. Tables of critical discrepancies between WMS-III and WAIS-III indices are presented in the technical manual. This will be of considerable potential interest in view of the clinical problem of differentiating between organic and functional memory impairment.

A possible solution to the problem of evaluating impairment would of course be to establish a National testing programme whereby, for example, all eighteen year olds were obliged to subject themselves to intellectual assessment. Although intellectual ability has been shown to increase, leading to the requirement to re-standardise the Wechsler scales, it would still be possible to evaluate a person's relative position with

respect to the contemporary normal distribution if retesting were required to assess potential impairment. Such a programme is, naturally, unlikely to be established, but it would be possible to justify it on the grounds that brain injury, for example, is extremely common. Jennet, Murray, Carlin, McKean, MacMillan & Strang (1979), reporting the Scottish Head Injuries Management Study showed that around one in sixty of the Scottish population present at hospital casualty departments each year having sustained concussional brain injuries. Teasdale (1997), showed that of 3005 patients admitted to one of five Glasgow hospitals in a one-year period, 91% had sustained a mild injury on the basis of severity of injury assessed on admission to hospital. In spite of this, 51% of the patients with mild injury were moderately or severely disabled at one year post injury. This was an unusual study which examined an entire geographical cohort irrespective of injury severity. Surprising numbers of less severely injured patients had significant post-injury adverse sequelae, and these cases are not often included in outcome studies. Social services contact was recorded in only 10% of cases in spite of the fact that over half were so severely disabled at a year that they were unable to work.

About one third of head injury victims are under 10 years of age and 54% are under 20 (Jennet *et al.,* 1979). Clearly, a significant number of cases would have been injured by the time they had presented for their National testing. However, given the annual incidence of concussional brain injury, and on the basis of average life expectancy of more than 60 years, each individual can expect to sustain at least one concussional brain injury in life. The value of premorbid data has been shown in a study

of intellectual impairment in ex-servicemen with multiple sclerosis (Canter, 1951) using data obtained on induction to the service.

Models which purport to be representative of the general population must be based on a carefully stratified normal sample, as was achieved in both this study, and in the US WAIS-R standardisation sample employed to generate the US demographic equations. The clinical sample did not correspond exactly with the UK healthy sample with regard to the years of education predictor variable. In fact, as noted, the clinical sample had significantly fewer years of education, and by implication would probably have had lower mean premorbid IQ. Such differences between the samples would have served to magnify the differentiation between the healthy and clinical cases.

The search for an adequate current ability measure which has good criterion validity and which is robust in the face of neurological disease goes on in view of the importance of quantifying intellectual deterioration in clinical practice. In particular, the CCRT (Beardsall & Huppert, 1994) and the Spot-The-Word Test (STW; Baddeley, Emslie & Nimmo-Smith, 1993) should theoretically be more resistant to impairment than NART. Preliminary work in head injury subjects has shown that performance on both measures did not, as previously noted, deteriorate in a sample in which NART performance was also maintained (Simons, 1997). This work was replicated in a study by Watt & O'Carroll (1999) in a clinical sample of 25 brain injury patients. These authors then examined the variance in current verbal intelligence accounted for by NART, CCRT and STW scores in a sample of 114 healthy adults. They found that both NART and CCRT accounted for approximately 50% of the variance in verbal IQ, and

when demographic information was included as predictor variables, there was improvement in the variance accounted for to 60% and 62% respectively. The findings in respect of STW were disappointing with evidence of relatively poor criterion validity: STW only accounted for 29% of the variance in verbal intelligence. Following addition of the demographic variables, this rose to only 41%. Conway & O'Carroll (1997) showed that in a sample of 30 patients with probable Alzheimer's disease CCRT performance was preserved relative to NART and that the former was not correlated with dementia severity unlike the NART.

What is now required is a study where the performance of these new present ability measures is examined in samples where NART performance has been shown to deteriorate. Ideal cases would be patients with Alzheimer's disease stratified for dementia severity, and patients with Huntington's Disease. In view of the evidence of relatively poor criterion validity of the STW (Watt & O'Carroll, 1999) a potentially useful research development would be to attempt to create a further reading task, theoretically as resistant to impairment as the STW, but with more acceptable criterion validity. One possibility to improve the relationship to verbal intelligence might be to increase the number of options from the present dichotomous arrangement whereby it is possible to guess correct 50% of the time. The more complex the task, however, the greater the burden on current ability and this would detract from the potential robustness of the task in impaired subjects.

As matters stand, the NART and CCRT appear to be the best current ability measures available to us, and it will be of some interest to see work examining their

relationship to the WAIS-III in addition to comparing their psychometric properties with the Wechsler Verbal Fluency Test when it is made available.

6.0 REFERENCES.

Anastasi, A. (1988). Psychological Testing, 6th ed. pp. 84-87. New York: Macmillan Publishing.

Atkinson, L. (1991). Some tables for statistically based interpretation of WAIS-R factor scores. Psychological Assessment,3, 288-291.

Atkinson, L., Cyr, J. J., Doxey, N. C. S. & Vigna, C. M. (1989). Generalizability of WAIS-R factor structure within and between populations. Journal of Clinical Psychology, 45, 124-129.

Baddeley, A., Emslie, H. & Nimmo-Smith, I. (1993). The Spot-the-Word Test: a robust estimate of verbal intelligence based on lexical decision. British Journal of Clinical Psychology, 32, 55-65.

Barona, A. & Chastain, R. L. (1986). An improved estimate of premorbid IQ for blacks and whites on the WAIS-R. International Journal of Clinical Neuropsychology, 8, 169-173.

Barona, A., Reynolds, C. R. & Chastain, R. (1984). A demographically based index of premorbid intelligence for the WAIS-R. Journal of Consulting and Clinical Psychology, 52, 885-887.

Beardsall, L. & Huppert, F. A. (1994). Improvement in NART word reading in demented and normal older persons using the Cambridge Contextual Reading Test. Journal of Clinical and Experimental Neuropsychology, 2, 232-242.

Blackmore, L. M., Crawford, J. R. & Simpson, S. A. (1994). Current and premorbid intelligence in Huntington's disease. Proceedings of the British Psychological Society,2, 57.

Blair, J. R. & Spreen, O. (1989). Predicting premorbid IQ: A revision of the National Adult Reading Test. The Clinical Neuropsychologist, 3, 129-136.

Borkowski, J.G., Benton, A. L. & Spreen, O. (1967). Word fluency and brain damage. Neuropsychologia, 5, 135-140.

Bornstein, R. A. (1983). VIQ-PIQ discrepancies on the WAIS-R in patients with unilateral or bilateral cerebral dysfunction. Journal of Consulting and Clinical Psychology, 51, 779-780.

Bornstein, R. A. & Matarazzo, J. D. (1982). Wechsler VIQ versus PIQ differences in cerebral dysfunction: A literature review with emphasis on sex differences. Journal of Clinical Neuropsychology, 4, 319-334.

Bornstein, R. A., Termeer, J., Longbrake, K., Heger, M. & North, R. (1989). WAIS-R cholinergic deficit profile in depression. Journal of Consulting and Clinical Psychology, 1, 342-344.

Brandt, J. (1991). Cognitive impairments in Huntington's Disease: insights into the neuropsychology of the striatum. In f. Boller & J. Grafman (Eds). Handbook of Neuropsychology, Volume 5, pp241-264. New York: Elsevier Science.

Brinkman, S. D. & Braun, P. (1984). Classification of dementia patients by a WAIS profile related to central cholinergic deficiencies. Journal of Clinical Neuropsychology, 6, 393-400.

Caine, E. D., Hunt, R. D., Weingartner, H. & Ebert, M. H. (1978). Huntington's dementia. Archives of General Psychiatry, 35, 377-384.

Canavan, A. G. M., Dunn, G., & McMillan, T. M. (1986). Principal components of the WAIS-R. British Journal of Clinical Psychology, 25, 81-85.

Canter, A. H. (1951). Direct and indirect measurement of deficit in multiple sclerosis: parts 1 & 2. Journal of General Psychology, 44, 3-25, 27-50.

Cohen, J. & Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioural sciences (2nd ed.). Hillsdale NJ: Erlbaum.

Conway, S. C. & O'Carroll, R. E. (1997). An evaluation of the Cambridge Contextual Reading Test (CCRT) in Alzheimer's disease. British Journal of Clinical Psychology, 36, 623-625.

Coolidge, F. L., Peters, B. M., Brown, R. E., Harsch, T. L. & Crookes, T. G. (1985). Validation of a WAIS-R algorithm for the early onset of dementia. Psychological Reports, 57, 1299-1302.

Crawford, J. R., Johnston, D. A., Mychalkiw, B. & Moore, J. W. (1997). WAIS-R performance following closed head injury: A comparison of the clinical utility of summary IQs, factor scores and subtest scatter indices. The Clinical Neuropsychologist, 11, 345-355.

Crawford, J. R. (1989). Estimation of premorbid intelligence: a review of recent developments. In J. R. Crawford & D. M. Parker (Eds.), Developments in clinical and experimental neuropsychology (pp. 55-74). New York: Plenum Press.

Crawford, J. R. (1990). The estimation of premorbid intelligence. Unpublished PhD thesis. University of Aberdeen.

Crawford, J. R. (1992). Current and premorbid intelligence measures in neuropsychological assessment. In J. R. Crawford, W. McKinlay & D. M. Parker (Eds.), A Handbook of neuropsychological assessment (pp. 21-49). London: Erlbaum.

Crawford, J. R. & Allan, K. M. (1996). WAIS-R Subtest Scatter: Base rate data from a healthy UK sample. British Journal of Clinical Psychology, 35, 235-247.

Crawford, J. R. & Allan, K. M. (1997). Estimating premorbid IQ with demographic variables: regression equations derived from a UK sample. The Clinical Neuropsychologist, 11, 192-197.

Crawford, J. R., Allan, K. M., Cochrane, R. H. B. & Parker, D. M. (1990). Assessing the validity of NART estimated premorbid IQs in the individual case. British Journal of Clinical Psychology, 29, 435-436.

Crawford, J. R., Allan, K. M., McGeorge, P. & Kelly, S. M. (1997). Base rate data on the abnormality of subtest scatter for WAIS-R short-forms. British Journal of Clinical Psychology, 36, 433-444.

Crawford, J. R., Allan, K. M., Stephen, D. W., Parker, D. M. & Besson, J. A. O. (1989). The Wechsler Adult Intelligence Scale - Revised (WAIS-R): factor structure in a UK sample. Personality and Individual Differences, 10, 1209-1212.

Crawford, J. R., Besson, J. A. O. & Ebmeier, K. P. (1990). Dementia: the role of neurotransmitter deficits. In T G Dinan (Ed.), Principles and practice of biological psychiatry (pp. 105-133). London: Clinical Neuroscience Publishers.

Crawford, J. R., Besson, J. A. O., Bremner, M., Ebmeier, K. P., Cochrane, R. H. B. & Kirkwood, K. (1992). Estimation of premorbid intelligence in schizophrenia. British Journal of Psychiatry, 161, 69-74.

Crawford, J. R., Besson, J. A. O. & Parker, D. M., (1988). Estimation of premorbid intelligence in organic conditions. British Journal of Psychiatry, 153, 178-181.

Crawford, J. R., Besson, J. A. O., Parker, D. M., Sutherland, K. M. & Keen, P. L. (1987). Estimation of premorbid intellectual status in depression (1987). British Journal of Clinical Psychology, 26, 313-314.

Crawford, J. R., Johnston, D. A., Michalkiw, B. & Moore, J. W. (1997). WAIS-R performance following closed head injury: A comparison of the clinical utility of summary IQs, factor scores and subtest scatter indices. The Clinical Neuropsychologist, 11, 345-355.

Crawford, J. R., Michalkiw, B., Johnston, D. A. & Moore, J. W. (1996). WAIS-R short forms: Criterion validity in healthy and clinical samples. British Journal of Clinical Psychology, 35, 638-640.

Crawford, J. R., Moore, J. W., & Cameron, I. M. (1992). Verbal fluency: a NART-based equation for the estimation of premorbid performance. British Journal of Clinical Psychology, 31, 327-329.

Crawford, J. R., Obonsawin, M. C. & Allan, K. M. (1998). PASAT and components of WAIS-R performance: Convergent and discriminant validity. Neuropsychological Rehabilitation, 8, 255-272.

Crawford, J. R., O'Carroll, R .E. & Venneri, A. (1998). Neuropsychological assessment of the elderly. In A.S.Bellack & M. Hersen (Eds), Comprehensive Clinical Psychology, Volume 7: Clinical Geropsychology (pp 133-169). Oxford: Pergamon.

Crawford, J. R., Stewart, L. E., Cochrane, R. Foulds, J., Besson, J. A. O., & Parker, D. M. (1989). Estimating premorbid IQ from demographic variables: a regression equation derived from a UK sample. British Journal of Clinical Psychology, 28, 275-278.

Crawford, J. R., Stewart, L. E., Cochrane, R., Parker, D. M. & Besson, J. A. O. (1989). Construct validity of the National Adult Reading Test: A factor analytic study. Personality and Individual Differences, 10, 585-587.

Crawford, J. R., Stewart, L. E., Parker, D. M., Besson, J. A. O., & Cochrane, R. (1989). Estimation of premorbid intelligence: combining psychometric and demographic approaches improves predictive accuracy. Personality and Individual Differences, 10, 793-796

Crawford, J. R., Stewart, L. E., Parker, D. M., Besson, J. A. O., & De Lacey, G. (1989). Prediction of WAIS IQ with the National Adult Reading Test: cross-validation and extension. British Journal of Clinical Psychology, 28, 267-273.

Cummings, J. L., Houlihan, J. P. & Hill, M. A. (1986). The pattern of reading deterioration in dementia of the Alzheimer type: Observations and implications. Brain and Language, 29, 315-323.

Ebmeier, K. P., Booker, K., Cull, A., Gregor, A., Goodwin, G. M. & O'Carroll, R. E. (1993). The validity of the National Adult Reading Test in estimating premorbid intellectual ability in long-term survivors of hemispheric glioma and whole brain irradiation - a pilot study. Psycho-Oncology, 2, 133-137.

Franzen, M. D., Burgess, E. J. & Smith-Seemiller, L. (1997). Methods of estimating premorbid functioning. Archives of Clinical Neuropsychology, 12, 711-738.

Fromm, D., Holland, A.L., Nebes, R.D. & Oakley, M.A. (1991). A longitudinal study of word reading ability in Alzheimer's Disease: Evidence from the National Adult Reading Test. Cortex, 27, 367-376.

Fuld, P. A. (1982). Behavioural signs of cholinergic deficiency in Alzheimer dementia. Ageing, 19, 193-196.

Fuld, P. A. (1984). Test profile of cholinergic dysfunction and of Alzheimer-type dementia. Journal of Clinical Neuropsychology, 6, 380-392.

Goldman, R. S., Axelrod, B. N., Tandon, R. & Berent, S. (1993). Spurious WAIS-R cholinergic profiles in schizophrenia. The Clinical Neuropsychologist, 7, 171-178.

Gronwall, D. (1977). Paced auditory serial-addition task: a measure of recovery from concussion. Perceptual and Motor Skills, 44, 367-373.

Hachinski, V. C., Linette, D., Zilhla, E., Du Boulay, G. H., McAllister, V. L., Marshall, J., Ross Russell, R. W., & Symon, L. (1975). Cerebral blood flow in dementia. Archives of Neurology, 32, 632-637

Hart, S., Smith, C. M. & Swash, M. (1986). Assessing intellectual deterioration. British Journal of Clinical Psychology, 25, 119-124.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G. & Curtiss, G. (1993). Wisconsin Card Sorting Test Manual: Revised and Expanded. Odessa, FL: Psychological Assessment Resources, Inc.

Heinrichs, R. W. & Celinski, M. J. (1987). Frequency of occurrence of a WAIS dementia profile in male head trauma patients. Journal of Clinical and Experimental Neuropsychology, 9, 187-190.

Hoehn, M. M. & Yahr, M. D. (1967). Parkinsonism: Onset, progression and mortality. Neurology, 17, 427-442.

Ivnic, R. J., Malec, J. F., Smith, G. E., Tangalos, E. G., Peterson, R. C., Kokmen, E. & Kurland, L. T. (1992). Mayo's older Americans normative studies: WAIS-R norms for ages 56 to 97. The Clinical Neuropsychologist, 6, 1-30.

Javoy-Agid, F. & Agid, Y. (1980). Is the meso-cortical dopaminergic system involved in Parkinson's disease? Neurology, 30, 1326-1330.

Jennet, B., Murray, A., Carlin, J., McKean, M., MacMillan, R. & Strang, I. (1979). Scottish head injuries management study: Head injuries in three Scottish neurosurgical units. British Medical Journal,2, 955-958.

Kareken, D. A. (1997). Judgement pitfalls in estimating premorbid intellectual function. Archives of Clinical Neuropsychology, 12, 701-709.

Kareken, D. A. & Williams, J. M. (1994). Human judgement and estimation of premorbid intellectual function. Psychological Assessment, 6, 83-91.

Kendrick, D. C., Gibson, A. J. & Moyes, I. C. A. (1979). The revised Kendrick Battery: Clinical studies. British Journal of Social andClinical Psychology, 18, 329-339.

Knight, R. G. & Godfrey, H. P. D. (1984). Assessing the significance of differences between subtests on the Wechsler Adult Intelligence Scale - Revised. Journal of Clinical Psychology, 40, 808-810.

Krull, K. R., Scott, J. G. & Sherer, M. (1995). Estimation of premorbid intelligence from combined performance and demographic variables. The Clinical Neuropsychologist, 9, 83-88.

Lawson, J. S., Inglis, J. & Stroud, T. W. F. (1983). A laterality index of cognitive impairment derived from a principal-components analysis of the WAIS-R. Journal of Consulting and Clinical Psychology, 51, 841-847.

Lea, M. (1986). A British supplement to the manual of the Wechsler Adult Intelligence Scale-Revised. San Antonio: Psychological Corporation.

Leckliter, I. N., Matarazzo, J. D., & Silverstein, A. B. (1986). A literature review of factor analytic studies of the WAIS-R. Journal of Clinical Psychology, 42, 332-342.

Levin, B. E., Llabre, M. M. & Weiner, W. J. (1989). Cognitive impairments associated with early Parkinson's disease. Neurology, 39, 557-561.

Lezak, M. D. (1983). Neuropsychological assessment (2nd ed). New York; Oxford University Press.

Lezak, M. D. (1988a). IQ: R.I.P. Journal of Clinical & Experimental Neuropsychology, 10, 351-361.

Lezak, M. D. (1988b). Neurological tests and assessment techniques. In F. Boller & J. Grafman (Eds). Handbook of Neuropsychology, Volume 1, pp 47-68. Elsevier.

Lezak, M. D. (1995). Neuropsychological assessment, 3rd ed. New York; Oxford University Press.

Lishman, W. A. (1987). Organic Psychiatry, 2nd ed. New York: Oxford University Press.

Logsdon, R. G., Teri, L., Williams, D. E., Vitiello, M V. & Prinz, P. N. (1989). The WAIS-R profile: a diagnostic tool for Alzheimer's disease? Journal of Clinical and Experimental Neuropsychology, 11, 892-898.

Massman, P. J. & Bigler, E. D. (1993). A quantitative review of the diagnostic utility of the WAIS-R Fuld profile. Archives of Clinical Neuropsychology, 8, 417-428.

Matarazzo, J. D., Daniel, M. H., Prifitera, A. & Herman, D.O. (1988). Inter-subtest scatter in the WAIS-R standardisation sample. Journal of Clinical Psychology, 44, 940-950.

Matarazzo, J. D. & Herman, D. O. (1985). Clinical uses of the WAIS-R: Base rates of differences between VIQ and PIQ in the WAIS-R standardisation sample. In B. Wolman (Ed.), Handbook of Intelligence: Theories, Measurement and Applications (pp. 899-932). New York: Wiley.

Matarazzo, J. D. & Prifitera, A. (1989). Subtest scatter and premorbid intelligence: Lessons from the WAIS-R standardization sample. Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1, 186-191.

Matarazzo, J. D. (1972). Wechsler's measurement and appraisal of adult intelligence (5th ed.). Baltimore: Williams & Wilkins.

McKinlay, W. W., Gray, J. M. (1992). Assessment of the severely head-injured. In J.R. Crawford, W. McKinlay & D.M. Parker (Eds.), A Handbook of Neuropsychological Assessment (pp 363-378. London, Erlbaum

McLean, J. E., Reynolds, C. R. & Kaufman, A. S. (1990). WAIS-R subtest scatter using the profile variability index. Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2, 289-292.

Miller, E. (1984). Verbal fluency as a function of a measure of verbal intelligence and in relation to different types of cerebral pathology. British Journal of Clilical Psychology, 23, 53-57.

Mortensen, E. L., Gade, A. & Reinisch, J. M. (1991). A critical note on Lezak's 'best performance method' in clinical neuropsychology. Journal of Clinical and Experimental Neuropsychology, 13, 361-371.

Nebes, R. D., Martin, D. C. & Horn, I. C. (1984). Sparing of semantic memory in Alzheimer's disease. Journal of Abnormal Psychology, 93, 321-330.

Nelson, H. E. (1982). The National Adult Reading Test. Test Manual. Windsor: NFER-Nelson.

Nelson, H. E. & McKenna, P. (1975). The use of current reading ability in the assessment of dementia. British Journal of Social and Clinical Psychology, 14, 259-267.

Nelson, H. E. & O'Connell, A. (1978). Dementia: The estimation of premorbid intelligence levels using the new adult reading test. Cortex, 14, 234-244.

Nelson, H. E. & Willison, J. (1982). National Adult Reading Test Manual (2nd ed.). Windsor: NFER-Nelson.

Obonsawin, M. C., Robertson, A., Crawford, J. R., Perera, C., Walker, S., Blackmore, L., Parker, D. M. & Besson, J. A. O. (1998). Non-mnestic cognitive function in the scopolamine model of Alzheimer's disease. Human Psychopharmacology, 13, 439-449.

O'Carroll, R. E. (1995). The assessment of premorbid ability. Neurocase, 1, 83-89.

O'Carroll, R. E. & Gilleard, C. J. (1986). Estimation of premorbid intelligence in dementia. British Journal of Clinical Psychology, 25, 157-158.

O'Carroll, R. E. (1987). The inter-rater reliability of the National Adult Reading Test (NART): A pilot study. British Journal of Clinical Psychology, 26, 229-230.

O'Carroll, R. E., Baikie, E. M. & Whittick, J. E. (1987). Does the National Adult Reading Test hold in dementia? British Journal of Clinical Psychology, 26, 315-316.

O'Carroll, R.E., Curran, S. M., Ross, M., Murray, C., Riddle, W., Moffoot, A. P. R., Ebmeier, K. P. & Goodwin, G. M. (1994). The differentiation of major depression from dementia of the Alzheimer type using within-subject neuropsychological discrepancy analysis. British Journal of Clinical Psychology, 33, 23-32.

O'Carroll, R. E., Moffoot, A., Ebmeier, K. P. & Goodwin, G. M. (1992). Estimating pre-morbid intellectual ability in the Alcoholic Korsakoff Syndrome. Psychological Medicine, 22, 903-909.

O'Carroll, R. E., Walker, M., Dunan, J., Murray, C., Blackwood, D., Ebmeier, K. P. & Goodwin, G. M. (1992). Selecting controls for schizophrenia research studies: The use of the National Adult Reading Test (NART) as a measure of premorbid ability. Schizophrenia Research, 8, 137-141.

Office of Population, Censuses and Surveys (1980). Classification of Occupations. London: HMSO.

Paolo, A. M., Ryan, J. J., Troster, A. I. & Hilmer, C. D. (1996). Demographically based regression equations to estimate WAIS-R subtest scores. The Clinical Neuropsychologist, 10, 130-140.

Patterson, K., Graham, N. & Hodges, J. R. (1994). Reading in dementia: A preserved ability? Neuropsychology,8, 395-407.

Quinn, N., Brown, R. & Marsden, C. D. (1986). Dementia and Parkinson's disease; a review of the current evidence. In P. E. Bebbington & R. Jacoby (Eds.), Psychiatric Disorders in the Elderly (pp. 25-53). Witham: Mental Health Foundation.

Raven, J. C. (1982). Revised Manual for Raven's Progressive Matrices and Vocabulary Scale. Windsor: NFER-Nelson.

Ryan, J. J., Paolo, A. M. & Brungardt, T. M. (1990). Standardization of the Wechsler Adult Intelligence Scale - Revised for persons 75 years and older. Psychological Assessment, 2, 404-411.

Ryan, J. J., Paolo, A. M. & Van Fleet, J. N. (1994). Neurodiagnostic implications of abnormal verbal-performance IQ descrepancies on the WAIS-R: a comparison with the standardizations sample. Archives of Clinical Neuropsychology, 9, 251-258.

Scatton, B., Rouquier, L., Javoy-Agid, F. & Agid, Y. (1982). Dopamine deficiency in the cerebral cortex in Parkinson's disease. Neurology, 32, 1039-1040.

Schlosser, D. & Ivison, D. (1989). Assessing memory deterioration with the Wechsler Memory Scale, the National Adult Reading Test, and the Schonell Graded Word Reading Test. Journal of Clinical and Experimental Neuropsychology, 11, 785-792.

Schmand, B., Geerlings, M. I., Jonker, C. & Lindeboom, J. (1998). Reading ability as an estimator of premorbid intelligence: Does it remain stable in emergent dementia? Journal of Clinical and Experimental Neuropsychology, 20, 42-51.

Schonell, F. J. (1942). Backwardness in the Basic Subjects. Edinburgh: Oliver & Boyd.

Sharpe, K. & O'Carroll, R. E. (1991). Estimating premorbid intellectual level in dementia using the National Adult Reading Test: a Canadian study. British Journal of Clinical Psychology, 30, 381-384.

Silverstein, A. B. (1981). Reliability and abnormality of test score differences. Journal of Clinical Psychology, 37, 392-394.

Silverstein, A. B. (1982). Pattern analysis as simultaneous statistical inference. Journal of Consulting and Clinical Psychology, 50, 234-240.

Silverstein, A. B. (1984). Pattern analysis: The question of abnormality. Journal of Consulting and Clinical Psychology, 52, 936-939.

Simons, J. (1997). The estimation of premorbid intelligence: A comparison of measures. Unpublished BSc thesis, University of Aberdeen.

Starkstein, S. E., Brandt, J., Folstein, S., Strauss, M., Berthier, M. L., Pearlson, G. D., Wong, D., McDonnell, A. & Folstein, M. (1988). Neuropsychological and neuroradiological correlates in Huntington's disease. Journal of Neurology, Neurosurgery, and Psychiatry, 51, 1259-1263.

Stebbins, G. T., Gilley, D. W., Wilson, R. S., Bernard, B. A. & Fox, J. H. (1990). Effects of language disturbances on premorbid estimates of IQ in mild dementia. The Clinical Neuropsychologist, 4, 64-68.

Taylor, K. I., Salmon, D. P., Rice, V. A., Bondi, M. W., Hill, R. L., Ernesto, C. R. & Butters, N. (1996). Longitudinal examination of American National Adult Reading Test (AMNART) performance in dementia of the Alzheimer type (DAT): Validation and correction based on degree of cognitive decline. Journal of Clinical and Experimental Neuropsychology, 18, 883-891.

Teasdale, G. (1997). Hospitalised concussional brain injury: rehabilitation and outcome at one year. Study No. K/OPR/2/2/D229, Chief Scientist's Office, The Scottish Office.

Tellegen, A. & Briggs, P. F. (1967). Old wine in new skins: Grouping Wechsler subtests into new scales. Journal of Consulting and Clinical Psychology, 31, 499-506.

Tomlinson, B. E., Irving, D. & Blessed, G. (1981). Cell loss in the locus coeruleus in senile dementia of Alzheimer type. Journal of Neurological Sciences, 11, 419-428.

Tulsky, D., Zhu, J. & Ledbetter, M. F. (1997). WAIS-III - WMS-III Technical Manual. San Antonio: Psychological Corporation.

van den Broek, M. D. & Bradshaw, C. M. (1994). Detection of acquired deficits in general intelligence using the National Adult Reading Test and Raven's Standard Progressive Matrices. British Journal of Clinical Psychology, 33, 509-515.

Vanderploeg, R. D. & Schinka J. A. (1995). Predicting WAIS-R IQ premorbid ability: combining subtest performance and demographic veriable predictors. Archives of Clinical Neuropsychology, 10, 225-240.

Vernon, P. A. (1983). Speed of information processing and general intelligence. Intelligence, 7, 53-70.

Walton, J. N. (1977). Brain's Diseases of the Nervous System, 8th ed. Oxford: Oxford University Press.

Watt, K. J. & O'Carroll, R. E. (1999). Evaluating methods for estimating premorbid intellectual ability in closed head injury. Journal of Neurology, Neurosurgery and Psychiatry,66, 474-479.

Webster, D. D. (1968). Critical analysis of the disability in Parkinson's Disease. Modern Treatment,5, 257-282.

Wechsler, D. (1958).The Measurement and Appraisal of Adult Intelligence. Baltimore: Williams & Wilkins.

Wechsler, D. (1981). Manual for the Wechsler Adult Intelligence Scale-Revised. New York: Psychological Corporation

Wechsler, D. (1987). Manual for the Wechsler Memory Scale-Revised. San Antonio: Psychological Corporation.

Wechsler, D. (1997a). Wechsler Adult Intelligence Scale-Third Edition. San Antonio: Psychological Corporation.

Wechsler, D. (1997b). Wechsler Memory Scale-Third Edition. San Antonio: Psychological Corporation.

Willshire, D., Kinsella, G. & Prior, M. (1991). Estimating WAIS-R IQ from the National Adult Reading Test: A cross-validation. Journal of Clinical and Experimental Neuropsychology, 13, 204-216.

Wilson, R. S., Rosenbaum, G., Brown, G., Rourke, D., Whitman, D. & Grisell, J. (1978). An index of premorbid intelligence. Journal of Consulting and Clinical Psychology, 46, 1554-1555.

Yates, A. (1956). The use of vocabulary in the measurement of intellectual deterioration. Journal of Mental Science, 102, 409-440.